ROUTING AND CONGESTION CONTROL STRATEGIES IN OPPORTUNISTIC NETWORKS: A SURVEY

Vidya K[1], Mr Hemanth S R[2]
[1]M.Tech, [2]Assistant Professor, Department of Computer Science
Maharaja Institute of Technology Mysore
Mysore, India

Abstract- Opportunistic networks are a class of mobile ad hoc networks (MANETs) where contacts between mobile nodes occur unpredictably and where a complete end-to-end path between source and destination rarely exists at one time. Due to mobility of nodes network topology regularly changes and so finding a delivery path to a destination is a challenging task. Two important functions, traditionally provided by the transport layer, are ensuring the reliability of data transmission between source and destination, and ensuring that the network does not become congested with traffic. However, modified versions of TCP that have been proposed to support these functions in MANETs are ineffective in opportunistic networks. Therefore, in order to make communication possible in an opportunistic network, the intermediate nodes may take custody of data during the blackout and forward it when the connectivity resumes. Routes are built dynamically, while messages are en route between the sender and the destination, and any possible node can opportunistically be used as next hop, provided it is likely to bring the message closer to the final destination. These requirements make opportunistic networks a challenging and promising research field. In this paper, we discuss some routing and congestion control strategies in opportunistic networks.

Keywords - MANETs, Intermittently connected networks, Opportunistic network, Routing strategy, Congestion control strategy.

I. INTRODUCTION

Mobile ad hoc networks (MANETs) are infrastructure less networks where mobile nodes can move freely. One node can directly communicate with another if they are within radio communication range. A node can simultaneously serve both as a source or destination of a message and as a relay for other messages. A message traverses the network being relayed from one node to another node until it reaches its destination (multi-hop communication). Since the nodes are moving, the network topology regularly changes and so finding a delivery path to a destination is a challenging task. Constructing end-to-end delivery paths and ensuring robust message delivery in the face of dynamic topology changes are challenges that have been addressed in MANETs and an abundance of routing and transport protocols have been proposed. In all these protocols, it is implicitly assumed that the network is continuously connected and that there exists at all times end-to-end paths between all source and destination pairs in the networks. However, in some scenarios complete end-to-end paths rarely or never exist between sources and destinations within the MANETs, due to high node mobility or low node density. These networks may experience frequent partitioning, with the disconnections lasting for long periods.

As a consequence, the end-to-end transfer delays in these intermittently connected networks (ICNs) are much greater than typical IP data transfer delays in conventional networks such as the Internet. Within ICNs we can identify opportunistic networks, which are networks where contacts between mobile nodes occur unpredictably because the node’s movement is effectively random, and where the duration of each node contact is also unpredictable. The challenges of developing efficient algorithms for opportunistic networks are different from those of classic ICNs.

A. ICN Overview

ICNs occur in challenged network environments examples include deep space communications where links have very long delays sparse sensor networks where connectivity is frequently intermittent , animal wildlife monitoring networks where animal movements are unpredictable, e.g. Zebranet, and in human (social) networks where connectivity occurs opportunistically, e.g. pocket-switched networks. In general, ICNs do not satisfy traditional networking assumptions, where end-to-end paths always exist, and the networks have low propagation delays or round-trip times, low bit error rates, and high bandwidth.

As a result, communication protocols built for these conventional networks, e.g. the Internet and MANETs, are not able to handle data communication efficiently in ICNs. End-to-end communication using the TCP/IP protocol suite is ineffective against the impairments of ICNs. In the network layer, MANET routing protocols, such as OLSR, AODV and DSR, will drop packets if the destination cannot be found. In the transport layer, TCP variants for MANETs, such as TCP-EFLN, A-TCP, TCP Snoop and TCP-Bus, will also break down in ICNs. These protocols assume that the network is continuously connected, and they consider link disruptions, due to node mobility or link layer contention, as temporary and short-term events. TCP eventually fails in ICNs, since link disconnections occur frequently and the round trip delays are too long.

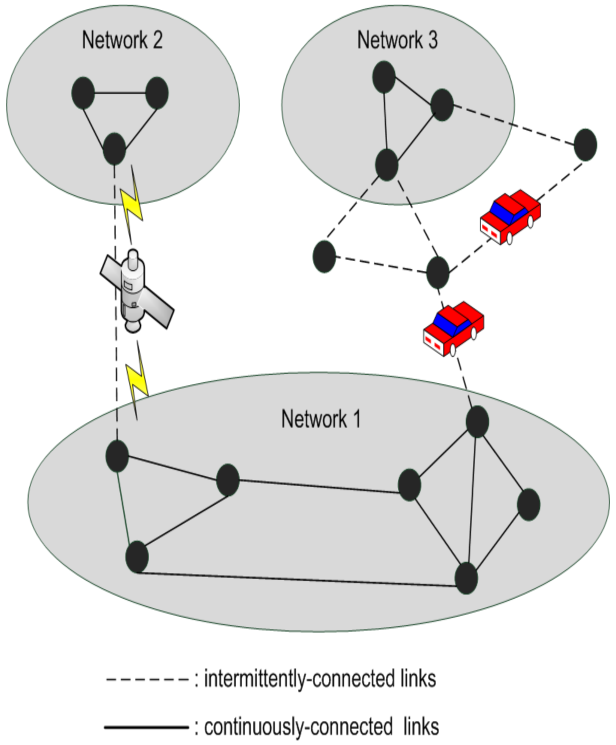
Network 2
Network 3
Network 1
------: intermittently-connected links
——: continuously-connected links

Figure 1: Intermittently connected networks

An example ICN scenario is illustrated in Figure 1, where three networks, each of which is continuously connected, are linked by intermittent connections, namely a satellite link (between networks 1 and 2) and a vehicular network (between networks 1 and 3). The satellite link is scheduled and predictable, whereas the vehicle-based links are unpredictable and therefore opportunistic. The vehicle contacts, when they occur, might be of long or short duration. ICN nodes (or simply "nodes" in this article) are responsible for managing data transfer between the temporarily disconnected networks. As nodes come into contact, they can transfer data, for example sending and receiving bundles. A bundle is an arbitrary sized data unit and has a time-to-live before bundle expiration; the term "message" is also used to refer to a "bundle". When a peer node or a link or path is currently not available, a node waits, storing the bundle forwarding it to another node that may have better a chance of delivering the bundle to its destination. Communications between disconnected areas can be performed by a store-forward (SF) mechanism, as in the satellite communications between network 1 and 2 or a store-carry-forward (SCF) mechanism, e.g. in the vehicular network between network 1 and 3. In SF, when there is no next hop known or no available link to the known next hop, bundles are stored in a node buffer waiting for the next contact event. In SCF, physical message carriers, such as vehicles, humans or message ferries, are added to carry and forward messages between disconnected areas For both mechanisms, the probability of node contact, the node contact duration and node resource capacity are key attributes for effective data delivery in ICN.

B. ICN Routing Strategies

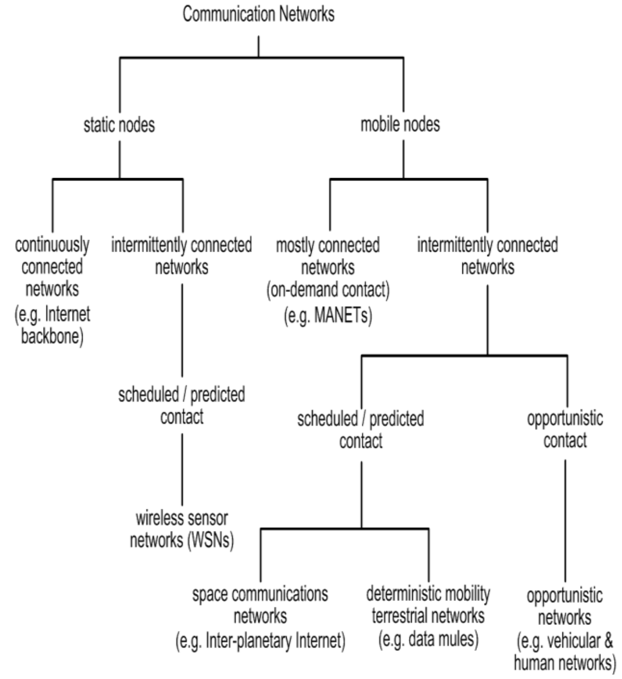
Communication Networks
static nodes
mobile nodes
continuously connected networks (e.g. Internet backbone)
intermittently connected networks
mostly connected networks (on-demand contact) (e.g. MANETs)
intermittently connected networks
scheduled / predicted contact
scheduled / predicted contact
opportunistic contact
wireless sensor networks (WSNs)
space communications networks (e.g. Inter-planetary Internet)
deterministic mobility terrestrial networks (e.g. data mules)
opportunistic networks (e.g. vehicular & human networks)

Figure 2: Taxonomy of communication networks

Routing in ICNs is more complicated than in MANETs due to the lack of up-to-date network topology information. Here the routing algorithms affect design decisions about transfer and congestion control mechanisms. ICN routing protocols typically use historical node contact data to predict future network topology. Three categories of regularity of node contacts can be defined, namely on-demand contact, scheduled or predicted contact and opportunistic contact as seen in Figure 2. Network is divided, based on node mobility, into static and mobile nodes. Static node networks can be either continuously connected (such as the Internet backbone) or intermittently connected. The latter division includes wireless sensor networks (WSNs), whose nodes conserve energy by disabling their radio connection when not required. In the mobile node branch of the taxonomy, we again distinguish between networks where links between nodes generally exist and networks where node contact is intermittent. In MANETs, links are assumed to be always or usually available when needed, this is also known as on-demand contact. We use the regularity of node contact to further divide the intermittently connected mobile networks. We distinguish between networks where node contacts are predicted (e.g. the Interplanetary Internet (IPN)) or scheduled (for example, data mules), and networks where node contacts are not generally predictable, such as vehicular networks and human networks. It is this latter category that is commonly called opportunistic networks. In scheduled/predicted contact, future node contacts are known in advance. Two examples of this are a link between an earth station and a satellite where the satellite's view schedule is known in advance, and a link between wireless sensor devices and a data mule, which visits a sensor device at regular times to collect data. In these cases, message transmissions can be scheduled in advance so that optimal delivery performance can be achieved. In Opportunistic meetings, a node knows nothing about future contacts or network topology. In this case a routing strategy can stochastically estimate future node contacts.

C. Poor performance of TCP in ICN's

The problem concerns TCP's reliable data transfer. This is implemented by a receiver returning an acknowledgement (ACK) to the source when messages are correctly received. In ICNs that have highly variable network delays, the message round trip time (RTT) cannot be calculated easily or used to set retransmission time-out (RTO) values. The source is therefore unable to detect a lost message promptly, it also has to keep the outstanding unacknowledged messages, potentially for a long time. Also, in order to maintain a reasonable throughput, TCP has to use a large window size, this is feasible for networks with reasonable delays (of the order of seconds) but not if the delay is of the order of hours or days. TCP has no explicit knowledge of the congestion state in networks. Instead, it implicitly couples the end-to-end transfer reliability and congestion control mechanisms through its acknowledgments. If the source receives three duplicate ACKs, or if TCP's retransmission timer expires, it assumes traffic congestion has occurred and it reduces the sending rate to limit the network congestion. This behavior does not work effectively in ICNs, which have frequent link disruptions and long transfer delays, an acknowledgement received by the source does not reflect the recent condition of the network and hence the source cannot respond to congestion accurately. Modified versions of TCP have been proposed for MANETs, for example TCP-EFLN, A-TCP and TCP Snoop. They are designed particularly to deal with wireless link disconnections due to node mobility or link layer contention, and assume that link disruptions are short-term events. During a link breakage, these TCP variants typically enter a standby state, freezing their parameters such as the congestion window and retransmission time-out values. When the link is re-established, TCP unfreezes the parameters and resumes the data transfer. In ICNs, however, where the link breaks may last for hours or days, the frozen TCP parameters are likely to be invalid for the resumed connections

II. OPPORTUNISTIC NETWORK OVERVIEW

We define an opportunistic network as one type of challenged networks where network contacts are intermittent or where link performance is highly variable or extreme. Contacts are intermittent, so an end-to-end path between the source and the destination may never exist. Therefore, TCP/IP protocol will break in this kind of environment because an end-to-end path between the source and the destination may only exist for a brief and unpredictable period of time. Long propagation and variable queuing delays might be introduced and many Internet protocols which are designed to assume quick return of acknowledgements and data can fail to work in such networks. In such a network, there does not exist a complete path from source to destination for most of the time. In addition, the path can be highly unstable and may change or break quickly. Therefore, in order to make communication possible in an opportunistic network, the intermediate nodes may take custody of data during the blackout and forward it when the connectivity resumes. In opportunistic networks the existence of a simultaneous path is not assumed to transmit a message between a sender and a receiver. Information about the context in which the users communicate is a key piece of knowledge to design efficient routing protocols in opportunistic networks. But this kind of information is not always available. When users are very isolated, context information cannot be distributed, and cannot be used for taking efficient routing decisions. One possible solution to resolve the above issues is to exploits node mobility and local forwarding in order to transfer data. Data can be stored and carried by taking advantage of node mobility and then forwarded during opportunistic contacts. Here entire chunks of message are transferred from one storage place to a storage place in another node along a path that is expected to reach the destination. Opportunistic networks are one of the most interesting evolutions of MANETs. In opportunistic networks, mobile nodes are enabled to communicate th each other even if a route connecting them never exists. Furthermore, nodes are not supposed to possess or acquire any knowledge about the network topology, which is instead necessary in traditional MANET routing protocols. Routes are built dynamically, while messages are en route between the sender and the destination and any possible node can opportunistically be used as next hop, provided it is likely to bring the message closer to the final destination. These requirements make opportunistic networks a challenging and promising research field. The applications of opportunistic network is typically used in an environment that is tolerant of long delay and high error rate. For example, Sami Network Connectivity (SNC) Project [1] focuses on establishing Internet communication for Sami population of reindeer herders who live in remote areas. In Zebranet [2], the researchers used a opportunistic network to track the wild zebras.

A. Challenges in Opportunistic networks

In an opportunistic network, when nodes move away or turn off their power to conserve energy, links may be disrupted or shut down periodically. These events result in intermittent connectivity. When there is no path existing between the source and the destination, the network partition occurs. Therefore, nodes need to communicate with each other via opportunistic contacts through store-carry-forward operation. In this section, we consider two specific challenges in an opportunistic network: the contact opportunity and the node storage.

1) Contact: Due to the node mobility or the dynamics of wireless channel, a node might make contact with other nodes at an unpredicted time. Since contacts between nodes are hardly predictable, they must be exploited opportunistically for exchanging messages between some nodes that can move between remote fragments of the network.

2) Storage constraint: To avoid dropping packets, the intermediate nodes should have enough storage to store all messages for an unpredictable period of time until next contact occurs. In other words, the required storage space increases a function of the number of messages in the network. Therefore, the routing and replication strategies must take the storage constraint into consideration.

B. Basic Opportunistic network scenario

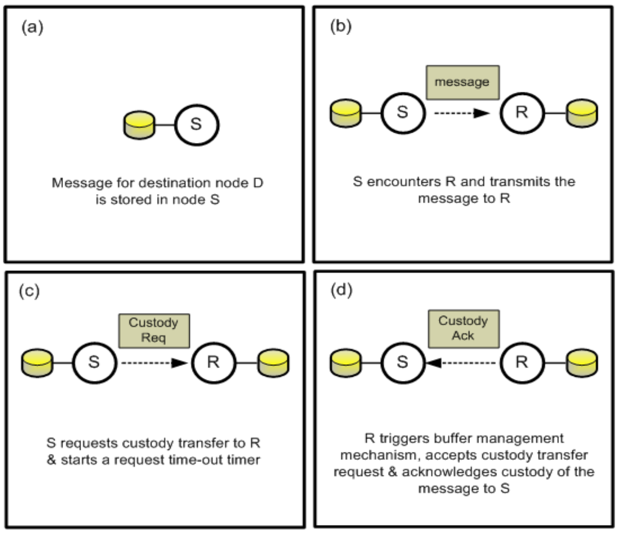
(a)
Message for destination node D is stored in node S
(b)
message
S encounters R and transmits the message to R
(c)
Custody Req
S requests custody transfer to R & starts a request time-out timer
(d)
Custody Ack
R triggers buffer management mechanism, accepts custody transfer request & acknowledges custody of the message to S
S
R

Figure 3: Opportunistic network scenario

We now describe a basic opportunistic network scenario and show how the transfer reliability and congestion control functions may interact. We consider the simple custody transfer scenario shown in Figure 3 A message destined for node D currently resides in the persistent storage of node S (Figure. 3(a)). During its travel, node S encounters node R and, based on its routing protocol, determines that node R is a better relay of the message to node D. Node S therefore forwards the message to R (Figure. 3(b)). S then requests a custody transfer service for the message to R and starts a request time-out timer (Figure. 3(c)). Upon receiving the custody request, R triggers its buffer management mechanism (part of the storage congestion control function) to determine whether receiving the message is likely to lead to buffer congestion in future, and therefore decides whether to accept or reject the custody request. In the example shown, R accepts the request (Figure .3(d)). In order to optimize the overall delivery success ratio, node buffer management needs to consider several attributes of a message, such as message priority, message lifetime, message size, and the probability of message being further forwarded. There are two forms of congestion in communication networks, namely link congestion and node storage congestion. A congested link occurs when two or more nodes that are within transmission range contend to transmit message using the same link or channel. However, congested links rarely occur in opportunistic networks. On the other hand, congested storage occurs when messages contend for the use of limited node storage space. In the remainder of this article, we will use the term “congestion” to refer to the “storage or buffer congestion” that more frequently occurs in opportunistic networks, given the (mobile) nodes’ limited storage capacity. Congestion control strategies in opportunistic networks are closely related to the number of message copies distributed throughout the network. Routing protocols may use a multiple copy strategy to increase the delivery ratio and/or to reduce end-to-end delivery latency. In this strategy, several copies of a message circulate in the network at any instant. Given the existence of redundant messages in the network it is likely that the provision of a custody service for messages is no longer needed, and in this case congestion control can be in the form of a message drop strategy. In the fixed Internet, packet dropping is typically performed in the network’s relay nodes, i.e. at IP routers. However, when an IP router drops messages during traffic congestion, it does not consider the overall delivery performance in the network. Instead, the end-to-end TCP mechanism ensures delivery, by requesting the source to retransmit the dropped messages.

In opportunistic networks, as we noted above, the long round trip time means that the end-to-end delivery mechanism is slow acting and hence dropped messages cannot be detected easily by the source. When an opportunistic network node has to drop messages during congestion, it needs to consider network delivery performance, for example by dropping those messages that have less impact on the end-to-end delivery. However, in the case of a single copy routing strategy, dropping messages during congestion may substantially decrease the overall delivery performance in the network. The congestion control strategy, or storage congestion management, should carefully select which messages are stored in a node so as to avoid future congestion. As an example, retaining messages that have longer remaining times to live (TTLs) is more risky and expensive for node buffer space than storing messages with small TTLs. TCP reduces its sending rate when it detects packet drops, as signaled by TCP’s acknowledgment mechanism. However, as we have noted this end-to-end approach is inappropriate in opportunistic networks. Instead, congestion control should be performed on per hop basis, and a node should use locally available congestion information to manage message flows. In Figure 4, we depict a typical node’s congestion-aware forwarding modules. The routing and congestion control modules work together to make forwarding decisions for messages in the buffer. During node contact, each module exchanges status data with its peer: the routing modules exchange routing information such as history contact data, delivery probability and node ranking, while the congestion control modules exchange node buffer statistics, for example buffer free space, queue growth rate, queuing delay and drop rate. A node will forward messages to a neighbor during contact if the neighbor meets the routing criteria and if the forwarded messages are unlikely to create congestion in the receiving neighbors buffer in the future.

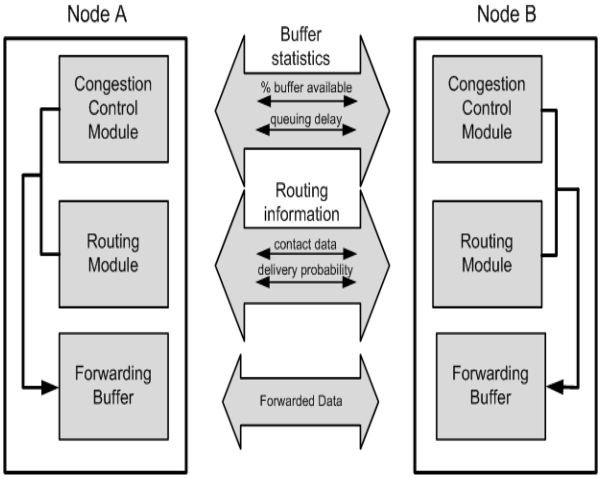

Figure 4: Congestion aware forwarding module in opportunistic network

III. ROUTING STRATEGIES IN OPPORTUNISTIC NETWORK

In this section, we discuss some routing solutions for an opportunistic network. Based on the number of copies of a message forwarded by the node, we can define two different routing schemes: forwarding-based (single copy) approach and flooding-based (multiple copies) approach. In the forwarding-based approach, there is only one single custodian for each message to help forwarding the message to destination. When the current custodian forwards the copy to an appropriate next-hop neighbor, this neighbors becomes the message's new custodian. The same process is repeated again and again until the message finally reaches its destination. This approach tries to reduce the buffer usages and the number of message transferred in the network. But it may suffer long delays and low delivery ratios. On the other hand, flooding-based approach may generate multiple copies of the same message. Each message can be routed independently for increased efficiency and robustness. This approach achieves lower delays and higher delivery ratio at the cost of a larger buffer space and more message transfers.

A. Forwarding-based approach

In the forwarding-based scheme, based on what type of knowledge nodes use to select the appropriate or the best path to destination node, the prior studies can be classified into three categories: direct-transmission, location-based and estimation-based.

1) Direct-transmission: Spyropoulos [3] proposed a simple single-copy routing called direct transmission routing. In this approach, after the source node generates a message, the message is hold by the source node until it reaches the destination node. The main advantage of this scheme is that it incurs minimum data transfers for message deliveries. On the other hand, although having minimal overhead, this scheme may incur very long delays for message delivery since the delivery delay for this scheme is unbounded [4].

2) Location-based: In the location-based approach, nodes will choose the neighbors who are closest to the destination to pass the message. LeBrun [5] proposed a method using the motion vector (MoVe) of mobile nodes to predict their future location. The MoVe scheme uses the knowledge of relative velocities of a node and its neighboring nodes to predict the closest distance between two nodes. After the nodes future location are calculated, messages are passed to nodes that are moving closer to the destination. As compared to epidemic routing, this approach has less control packet overhead and buffer usage. Leguay [6] presented a strategy that uses a virtual coordinate routing called mobility pattern spaces (MobySpace). The measure of closeness represents the probability that the nodes will come into contact with each other. They showed that this approach consumes less resources than epidemic routing.

3)Knowledge-based: In the knowledge-based approaches, based on certain knowledge about the network, the source and intermediate nodes decide which node to forward the messages as well as whether it should transmit the message immediately or hold the message until it meets a better node. Jain [7] proposed knowledge based routing scheme which is the first study in this area. Depending on the amount of knowledge about network topology characteristics and traffic demand, they define four knowledge oracles. Each oracle presents some particular knowledge of network. Based on the available oracles, the authors present a corresponding routing algorithm. The basic idea of their routing algorithms is to apply the traditional shortest path routing techniques to opportunistic network by exploiting the knowledge oracles. At the same time, author use the source routing to forward the message over the shortest path. This scheme formulates the routing in order to minimize the end-to-end delivery latency. Musolesi [8] present the Context-Aware Routing (CAR) protocol that provides an asynchronous communication for message delivery. In an opportunistic network, since the receiver is often not in the same connected network, synchronous delivery of messages is typically not possible. In CAR, if a message cannot be delivered synchronously, the message is sent to a host that has the highest probability of successful delivery and acts as a message carrier. The delivery probability process is based on the evaluation and prediction of context information using Kalman filters. The prediction process is used during temporary disconnection and the process is continued until it is possible to guarantee certain accuracy. Simulations shows that if the buffer size is small, the packet delivery ratio of CAR is better than that of epidemic routing due to that CAR only creates a single copy for each message. J. Burgess [9], proposes MaxProp protocol which is based on prioritizing both the schedule of packets transmitted to other peers and the schedule of packets to be dropped. These priorities are based on the path likelihoods to peers according to historical data and also on several complementary mechanisms, including acknowledgments, a head-start for new packets, and lists of previous intermediaries. MaxProp protocol, addresses scenarios in which either transfer duration or storage is a limited resource in the network. MaxProp address several problems that have observed in network topology. Existing approaches have a bias towards short-distance destinations, which MaxProp addresses by using hop counts in packets as a measure of network resource fairness. Additionally, existing approaches fail to remove stale data from network buffers. MaxProp uses acknowledgments that are propagated network wide, and not just to the source. Finally, MaxProp stores a list of previous intermediaries to prevent data from propagating twice to the same node. At the core of the MaxProp protocol is a ranked list of the peer's stored packets based on a cost assigned to each destination. The cost is an estimate of delivery likelihood. In addition, MaxProp uses acknowledgments sent to all peers to notify them of packet deliveries. MaxProp assigns a higher priority to new packets, and it also attempts to prevent reception of the same packet twice. When two peers discover each other, MaxProp exchanges packets in a specific priority order: First, all messages destined to the neighbor peer are transferred. Second, routing information is passed between peers, specifically, a vector listing estimations of the probability of meeting every other node. Third, acknowledgments of delivered data are transferred, regardless of source and destination.

An acknowledgment consists of the cryptographic hash of the content, source, and destination of each message, and is therefore about 128 bits. This mechanism serves to clear out buffers in the network of old data at little cost if the acknowledgment is small compared to data packets. Fourth, packets that have not traversed far in the network are given priority. MaxProp attempts to give new packets a head start in the network by placing them at a higher priority. The effect of this approach is that newer packets are transmitted at several transfer opportunities when they are first generated, increasing their chance of reaching the destination. Author has proposed MaxProp as an effective protocol for DTN routing, particularly for the context of our real DTN deployment. MaxProp unifies the problem of scheduling packets for transmission to other peers and determining which packets should be deleted when buffers are low on space. Additionally, have identified several complementary mechanisms for improving the performance of path-likelihood based routing, including: system-wide acknowledgments, hop lists denoting previous intermediate recipients, and priority for new packets using an adaptive threshold. Kun [10] proposed a shortest expected path routing (SEPR) similar to link-state routing to maintain a topology map to each other. SEPR first estimates the link forwarding probability based on history data. When two nodes meet, they exchange the link probability update messages called effective path length (EPL). A smaller EPL value suggests a higher probability of delivery. When a node received a smaller EPL, it will update its local EPL value. EPL is also used in deciding which nodes to forward the messages. Using SEPR protocol, the same message could be forwarded to multiple nodes to increase reliability and to reduce delay.

B. Flooding-based approach

In the flooding-based approach, every node broadcasts the received packet to all of its neighbors. However, in an intermittently connected network, some nodes might not be able to receive the broadcast packets due to network partitions. Therefore, each node stores the messages until the messages finally arrive the destination.

1) Epidemic routing: Epidemic routing is first proposed by Vahdat and Becker [11] for forwarding data in an opportunistic network. Epidemic routing utilizes the epidemic algorithm [12] that was originally proposed for synchronizing replicated databases. The epidemic algorithm ensures that a sufficient number of random exchanges of data in the network and guarantees all nodes will eventually receive all messages. The Epidemic Routing is similar to the flooding routing because it tries to send each message to all nodes in the network. Amin Vahdat proposes techniques to deliver messages in the case where there is never a connected path from source to destination or when a network partition exists at the time a message is originated. To this end, introduce Epidemic Routing, where random pair-wise exchanges of messages among mobile hosts ensure eventual message delivery. The goals of Epidemic Routing are to: i) maximize message delivery rate, ii) minimize message latency, and iii) minimize the total resources consumed in message delivery. The goal of this work is to develop techniques for delivering application data with high probability even when there is never a fully connected path between source and destination. Thus, with minimal assumptions about the connectivity of the underlying ad hoc network: i) the sender is never in range of any base stations, ii) the sender does not know where the receiver is currently located or the best "route" to follow, iii) the receiver may also be a roaming wireless host, and iv) pairs of hosts (not necessarily the sender and receiver) periodically and randomly come into communication range of one another through node mobility. Epidemic Routing is to distribute application messages to hosts, called carriers, within connected portions of ad hoc networks. In this way, messages are quickly distributed through connected portions of the network. Epidemic Routing then relies upon carriers coming into contact with another connected portion of the network through node mobility. At this point, the message spreads to an additional island of nodes. Through such transitive transmission of data, messages have a high probability of eventually reaching their destination. Author presented techniques to allow message delivery in the case where a connected path from source to destination is never available in mobile ad hoc networks. While existing ad hoc routing protocols are robust to rapidly changing network topology, they are unable to deliver packets in the presence of a network partition between source and destination. For a number of compelling application classes, including mobile sensor networks and disaster recovery scenarios, nodes can be spread over wide geographical distances. Such wide dispersion makes it unlikely that a connected path can always be discovered, making it virtually impossible to perform message delivery using current ad hoc routing protocols. Thus, Epidemic Routing, where random pair-wise exchanges of messages among mobile hosts ensure eventual message delivery. The goals of Epidemic Routing are to maximize message delivery rate and to minimize message latency while also minimizing the total resources consumed in message delivery. Epidemic Routing incurs significant demand on both bandwidth and buffer. To bound the overhead of delivering a message, Spyropoulos [13] proposed a technique called Spray and Wait to control the level of flooding. In the spray phase, there are L numbers of copies that are initially spread over the network by the source node or other nodes to L distinct relays. In the wait phase, if the destination was not found during the spray phase, each node who has a copy of message will perform direct transmission. Binary spray and wait is a variation of Spray and Wait and produces a better performance. In this approach, the binary spray source node send half of the copies of the message to the new relay node, and keeps the rest to itself. The source node and relay nodes uses repeat this procedure until there is only one copy left. When it is only one copy left, it switches to direct transmission.

2) Estimate/Prediction routing: In Estimate/ Prediction routing, nodes do not blindly forward the messages to all or some neighbors. Instead, nodes estimate the probability of each link to destination and use this information to decide whether it should store the packet and wait for a better chance as well as to decide which nodes to forward. A. Lindgren [14], presents PROPHET protocol, the random way-point mobility model is popular to use in evaluations of mobile ad hoc protocols, real users are not likely to move around randomly, but rather move in a predictable fashion based on repeating behavioral patterns such that if a node has visited a location several times before, it is likely that it will visit that location again. Author uses these observations and this information to improve routing performance by doing probabilistic routing and thus, propose PROPHET, a Probabilistic Routing Protocol using History of Encounters and Transitivity. To accomplish this, author establish a probabilistic metric called delivery predictability, P (a; b) 2 [0; 1], at every node *a* for each known destination *b*. This indicates how likely it is that this node will be able to deliver a message to that destination. The operation of PROPHET is similar to that of Epidemic Routing. When two nodes meet, they exchange summary vectors which in this case also contain the delivery predictability information stored at the nodes. This information is used to update the internal delivery predictability vector and then the information in the summary vector is used to decide which messages to request from the other node based on the forwarding strategy used.

E. Daly [15], proposes SimBet routing, message delivery in sparse Mobile Ad hoc Networks (MANETs) is difficult due to the fact that the network graph is rarely (if ever) connected. A key challenge is to find a route that can provide good delivery performance and low end-to-end delay in a disconnected network graph where nodes may move freely. This paper presents a multidisciplinary solution based on the consideration of the so called small world dynamics which have been proposed for economy and social studies and have recently revealed to be a successful approach to be exploited for characterizing information propagation in wireless networks. To this purpose, some bridge nodes are identified based on their centrality characteristics, i.e., on their capability to broker information exchange among otherwise disconnected nodes. Due to the complexity of the centrality metrics in populated networks the concept of ego networks is exploited where nodes are not required to exchange information about the entire network topology, but only locally available information is considered. Then SimBet Routing is proposed which exploits the exchange of pre-estimated ‘betweenness’ centrality metrics and locally determined social ‘similarity’ to the destination node. Author presents simulations using real trace data to demonstrate that SimBet Routing results in delivery performance close to Epidemic Routing but with significantly reduced overhead. Additionally, shows that SimBet Routing outperforms PRoPHET routing, particularly when the sending and receiving nodes have low connectivity.

IV. CONGESTION CONTROL STRATEGIES IN OPPORTUNISTIC NETWORKS

Kevin Fall [16], proposes the custody transfer mechanism proposed for enhancing reliability in delay-tolerant networks. This mechanism, which utilizes hop-by-hop transfer of reliable delivery responsibility, shares many features in common with a database transaction. A delay tolerant network consists of a directed graph G = (E, V) where the set of directed edges E are derived from a list of contacts. A contact describes a link's tail and head vertex, existence interval, plus its capacity and latency during the interval. An edge e = (t; h) is placed in the set E if t and h ever appear in a contact. The set of vertices V consist of store-and-forward message routers which may optionally provide custody transfer. Accepting a message with custody transfer amounts to promising not to delete it until it can be reliably delivered to another node providing custody transfer, to the best of the ability of the forwarder. Nodes holding a message with custody are called custodians. Ordinarily, there is a single custodian for a message, but in some circumstances more than one custodian owns a message or message fragment .Applications optionally request custody transfer to be performed on a per-message basis, and are delivered a custody acknowledgment when their host system has been able to move the message to one or more other nodes that are willing to take custody for it. In particular, the custody acknowledgment is not an end-to-end acknowledgment, but instead indicates that the responsibility for end-to-end reliable delivery has been delegated to some other party apart from the sending node. The custody transfer mechanism proposed for delay tolerant networks, with particular emphasis on its implications for congestion management and the semantics of its protocol operation. The congestion management problem can lead to a form of head-of-line blocking, and several techniques are available to handle the problem.

M. Radenkovic [17], propose congestion aware opportunistic forwarding that supports optimization of high volume multipoint data flows transfer while maintaining high buffer availability. Author propose and investigate new metric for analyzing and integrating node buffer and delay behavior with node's ego network buffer and delay behavior in a number of new heuristics for different forwarding strategies. The new forwarding strategies aim to allow avoiding the nodes and the parts of the network with high congesting rates with the aim to keep high success ratio, low delays and good network efficiency even at times of increasing congestion.

In particular, design combines, social driven part that aims to enable the most direct route to a destination node by selecting the intermediaries with higher probability of meeting the destination according to a social metric, node resources driven part that aims to detect and avoid the nodes that have low buffer availability, high delays or high congesting rate, ego network driven part that aims to detect and avoid parts of the networks that have low buffer and increased delay. In this way protocol works as a local forwarding protocol that diverts the load from its conventional social aware path at times of congestion and directs it via a different path that decreases the load of hotspots and end-to-end delays while keeping high success ratios. When a potential intermediary node or its ego network (contacts that it has "seen") is about to get increasingly congested, we determine the load and expected delays of a set of neighbors and their ego networks, and choose alternative targets for offloading the messages. Statistical analysis is performed for nodes contacts, storage and delay history in order to make a decision as to whether to offload messages to it, or not. Since CAF uses buffer statistics collected from nodes in the ego-network to calculate the node's local congestion level, an opportunity to improve the algorithm, by considering the structural properties of the neighboring nodes in the network. Since an ego network is the first-order neighborhood of a node (the ego), it only considers direct neighbors, and disregards the neighbors of the ego's neighbors. In highly clustered networks such as social networks, a node (or individual) that has neighboring nodes with high centralities tends itself to be more central as well, and therefore it is more likely to receive more traffic. By considering neighbors centralities, a node can improve the CAF local congestion calculation. J.M. Pujol [18] ,proposes Fair route like CAF is a proposal that addresses the unfair load distribution in social opportunistic networks. This forwarding strategy also relies on both routing and congestion control modules to make a forwarding decision during node contact. The routing module uses the perceived level of interaction with neighbor nodes to make routing decisions. This interaction level, or tie strength, represents the probability of a future contact between a pair of nodes. The tie strength increases with node contact events, but decreases exponentially over time. To achieve a balanced traffic distribution, node buffer statistics are also considered in the forwarding strategy. The congestion control module only considers the buffer queue length. The algorithm applies an assortative based queue control, where nodes will only accept a forwarding request from other nodes of equal or higher "status" (the term assortative is borrowed from sociology where people with similar social status tend to interact together, but disregard interactions with individuals of lower status). Here, node status is defined by the size of the node's queue length, with a longer queue length being a higher status. Thus, in Fair Route higher status nodes (nodes with longer buffer queue length) will be able to forward their messages faster, while lower status nodes will have to find alternative paths. As social opportunistic networks typically show a diversity of delivery paths between any two end nodes the authors then claim that the assortative-based congestion control does not necessarily imply a reduction of overall throughput, and that it has a positive impact on traffic distribution fairness in the network. T. Kathiravelu [19], introduced the Adaptive Routing protocol, which relies on a predictability metric that measures the degree of connectivity between a node and its neighbor. This favors more popular nodes (i.e. those having better connectivity) as relay nodes to increase the delivery likelihood of a message. However, since this strategy increases congestion in the most connected nodes, the authors later proposed their Congestion Aware Adaptive (CAA) algorithm to address the Adaptive Routing algorithm's drawback. The CAA algorithm improves a "naive" congestion control approach in which a node simply advertises its buffer free space to other nodes. Instead each node initially performs a self-assessment of its connectivity to its neighbors (a routing task) and then calculates a safety margin for its buffer according to its popularity level (a congestion control task). The buffer safety margin rises and falls with the increase or decrease respectively of the node's popularity level. In addition, the CAA algorithm favors receiving messages destined for the more popular nodes to reduce queue waiting time, hence reducing the probability of buffer congestion. At each node contact, the nodes exchange their storage availability information, i.e. buffer free space and threshold, as well as a list of nodes with the highest delivery predictability. A sending node is then allowed to forward a message to its contact node only if the message size satisfies the receiver's allowed buffer margin and the destination is in the receiver's list of node with high delivery probability. However, despite its simplicity we still see a potential drawback of the algorithm especially in large scale networks. Since, the algorithm requires every node to maintain a node predictability table, this table will grow linearly with the increasing number of nodes in the network, with a consequent scalability issue in large networks. Hua [20] argue that congestion occurrence in a custody node is a gradual procedure, and that early detection of congestion can be performed by assessing the node's state. They define three states, namely normal state (NS), congestion adjacent state (CAS) or congestion state (CS). The examination considers the rate at which node storage is used up. When the storage utilization exceeds a predefined level with most of the storage space used and the rate of increase of storage occupancy exceeds some threshold, the node is close to congestion and is defined as CAS. Then, if the storage utilization continues to increase and reaches another level with storage nearly exhausted and the rate of increase of the storage occupancy does not drop below the given threshold for a certain time interval, the node is congested and is marked as state CS. During node contacts, the node state is broadcast to all neighbors during opportunistic contacts, notifying them of the node's congestion status. When a node enters CAS, the neighbors mark the link to the node as partially congested, meaning that any paths that include the link should be avoided unless no other link is available. On the other hand, when a node is congested and in CS, the neighbors cannot choose a link to the node irrespective of network condition. Thus, the path avoidance algorithm refrains from forwarding a message to a node that is close to congestion or is actually congested. We can see that the effectiveness of the algorithm relies on how far the node congestion information can be broadcast. Ideally, the farther the information is propagated, the better the algorithm performs since k-hop neighboring nodes can redirect their traffic away from the congested node. However, as the broadcasting of information will be delayed in intermittently-connected networks, neighbors further away from the congested node may receive out-of-date node state information, leading them to choose inappropriate paths to message destinations. The trade-off between information broadcast range and overall delivery performance plays an important role in the algorithm and needs to be further investigated. Token Based Congestion Control (TBCC) [21] is a congestion avoidance proposal that attempts to match the volume of messages injected into a network with the total network capacity, i.e. the volume of messages the network can deliver to destinations in a bounded time. The algorithm is similar to Token Ring/Bus in that a node must possess a token to transmit data, but differs in that it only needs a token to inject a new message into the network. The algorithm views the network as a black box and the cost for a node to inject a single message into the black box is a single token (assuming a constant message length). A token can be reclaimed when a message leaves the network, i.e. when a message arrives at the destination or when the message's lifetime expires. TBCC furthermore assumes all nodes in the network cooperate in forwarding messages and are entitled to share available tokens equally. Tokens are initially evenly distributed among nodes in the network. When a source node wants to forward a message to a relay node, it initially checks its own token availability, transmits the message if its token count is greater than zero, and decrements its token count after successfully transmitting the message. If the token count is zero, the source node can query the peer node, asking for an extra token. Message transmission between relay nodes does not incur any token reduction, since tokens are only used when a message is initially injected into the network by the source. The authors' experiments assumed a constant number of nodes and tokens, which represent the total network capacity, and showed that the algorithm was able to manage message delivery and minimize node storage congestion probability. Despite the algorithm's simplicity, we consider that in practice the assumption is unrealistic in open networks, e.g. social opportunistic networks, since mobile users can autonomously join and leave the network at any time. Calculating the network capacity that corresponds to the number of tokens provided in the network is a challenging task if the number of active nodes in the network varies with time.

V. CONCLUSION AND FUTURE WORK

In this paper, an integrated routing for opportunistic networks has been proposed. We have observed that our proposed integrated routing is able to meet out the challenges of other routing schemes for the opportunistic networks, particularly the message delay and delivery probability, when context information about user is available or not. The present finding clearly indicates that the forwarding based is a very interesting approach of communication in opportunistic networks, however, in comparison to flooding-based protocols it is not suitable. The present routing is able to give better result in presence as well as absence of context information, specifically in term of message delay and delivery probability. Despite this, a number of directions exist in integrated routing which can be further investigated. For example we can improve performance of integrated routing in terms of message delay, message delivery, network congestion and resource consumption etc.

REFERENCES

[1] A. Lindgren and A. Doria, "Experiences from deploying a real life dtn system," In 2007 4th IEEE Consumer Communications and Networking Conference, pp. 217–221, 2007.

[2] "The zebranet wildlife tracker," http://www.princeton.edu/ mrm/zebranet.html.

[3] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Single-copy routing in intermittently connected mobile networks," In 2004 First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks, pp. 235 – 244, 2004. 1676

[4] M. Grossglauser and D. N. C. Tse, "Mobility increases the capacity of ad-hoc wireless networks," In IEEE/ACM Transactions on Networking, vol. 10, no. 4, pp. 477–486, 2002.

[5] J. LeBrun, C.-N. Chuah, D. Ghosal, and M. Zhang, "Knowledge based opportunistic forwarding in vehicular wireless ad hoc networks," In IEEE Vehicular Technology Conference (VTC), pp. 2289–2293, May 2005.

[6] J. Leguay, T. Friedman, and V. Conan, "Dtn routing in a mobility pattern space," In Proceeding of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking, pp. 276–283, 2005.

[7] S. Jain, K. Fall, and R. Patra, "Routing in a delay tolerant network," In Proceedings of ACM SIGCOMM, vol. 34, pp. 145–158, 2004.

[8] M. Musolesi, S. Hailes, and C. Mascolo, "Adaptive routing for intermittently connected mobile ad hoc networks," In IEEE WoWMoM 2005, pp. 183–189, 2005.

[9] J. Burgess, B. Gallagher, D. Jensen, and B. N. Levine, “Maxprop: Routing for vehicle-based disruption-tolerant networks,” In Proceedings of IEEE INFOCOM 2006, pp. 1–11, 2006.

[10] K. Tan, Q. Zhang, and W. Zhu, “Shortest path routing in partially connected ad hoc networks,” In Proceedings of IEEE Globecom 2003, vol. 2, pp. 1038–1042, 2003.

[11] A. Vahdat and D. Becker, “Epidemic routing for partially connected ad hoc networks,” In Duke University, Technical Report CS-200006, 2000.

[12] A. A. Demers, D. Greene, C. Hauser, W. Irish, J. Larson, S. Shenker, H. Sturgis, D. Swinehart, and D. Terry, “Epidemic algorithms for replicated database maintenance,” In Proceedings of the Sixth Symposium on Principles of Distributed Computing, pp. 1–12, 1987.

[13] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, “Spray and wait: An efficient routing scheme for intermittently connected mobile networks,” In Proceeding of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking WDTN ’05, pp. 252–259, 2005.

[14] A. Lindgren, A. Doria, O. Schelen, “Probabilistic Routing in Intermittently Connected Networks”, ACM SIGMOBILE Mobile Computing and Communication Review, vol.7, no. 3, pp. 19-20, July 2005.

[15] E. Daly, M. Haahr, “Social Network Analysis for Routing in Disconnected Delay Tolerant MANETs”, Proc. 8th ACM Intl. Sym. on Mobile Ad Hoc Networking and Computing, Montreal, Canada, Sep. 2007.

[16] K. Fall, W. Hong, S. Madden, “Custody Transfer for Reliable Delivery in Delay Tolerant Networks”, Technical Report IRB-TR-03-030, Intel Research Berkeley, 2005.

[17] M. Radenkovic, A. Grundy, “Congestion Aware Forwarding in Delay Tolerant and Socia Opportunistic Networks”, Proc. 8th Intl. Confon Wireless On-Demand Network Systems and Services, Bardonecchia,Italy, 2011.

[18] J. M. Pujol, A.L. Toledo, P. Rodriguez, “Fair Routing in Delay Tolerant Networks”, Proc. IEEE INFOCOM, Rio de Janeiro, Brazil, 2009.

[19] T. Kathiravelu, N. Ranasinghe, A. Pears, “Towards Designing a Routing Protocol for Opportunistic Network”, Proc. Intl. Conf. on Advances in ICT for Emerging Regions, Colombo, Sri Lanka, Sep. 2010.

[20] D. Hua, X. Du, L. Cao, G. Xu, Y. Qian, “A DTN Congestion Avoidance Strategy based on Path Avoidance”, Proc. 2nd Intl. Conf. Future Computer and Commun., Wuhan, China, 2010.

[21] E. Coe, C. Raghavendra, “Token Based Congestion Control for DTNs”, Proc. Aerospace Conference 2010, Big Sky, Montana, USA, 2010.