Expert Systems With Applications 57 (2016) 178–191

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A multi-objective optimization of data dissemination in delay tolerant networks

D. G Reina[a], R.I. Ciobanu[b], S.L. Toral[a,*], C. Dobre[b]

[a] Escuela Superior de Ingenieros, University of Seville, Avda. Camino de los Descubrimientos s/n, 41092 Sevilla, Spain
[b] Faculty of Automatic Controls and Computers, University Politehnica of Bucharest, Spl. Independentei 313, Bucharest, Romania

ARTICLE INFO

Article history:
Received 31 January 2015
Revised 22 March 2016
Accepted 23 March 2016
Available online 28 March 2016

Keywords:
Delay tolerant networks
Multi-objective optimization
Genetic algorithm
Data dissemination
Decision tree

ABSTRACT

Data dissemination in delay tolerant networks is an important issue due to the complexity of a multi-hop network formed by a high number of nodes with limited resources, variable and unpredicted mobility conditions. Nodes have to act as expert systems and make suitable forwarding decisions based on local knowledge on the fly. Most of the proposed algorithms rely on adjusting a range of decision variables related to social and topological aspects of the network. Adjusting such parameters is still an open issue since many of them are interrelated. To solve this problem, we propose a multi-objective evolutionary simulation framework for optimizing in terms of delivery hit, delivery cost and latency, a probabilistic data dissemination algorithm based on well-known and widely used social and topological parameters such as centrality, similarity, social strength, friendship, and trust. The proposed multi-objective based optimization framework provides many advantages with respect to existing approaches based on single objective optimization. Primarily, it allows the network designer to have a complete view of the possible outcomes of the data dissemination algorithm through the Pareto front (non-dominated solutions). Furthermore, we propose a decision tree-based selection to obtain under which values of the decision variables we can find a set of solutions that meet a target performance. We validate this selection mechanism by providing the conditions under which we can find balanced solutions in the considered simulation scenarios. The solutions provided by the proposed approach have significant implications for the design of new data dissemination algorithms in DTNs.

1. Introduction

Delay tolerant networks (DTNs) are decentralized networks where mobile nodes cooperatively communicate to transmit application data from a source node to a destination node. DTN topologies are dynamic, since nodes have a high degree of mobility, which leads to network partitioning, disconnections between nodes, and high end-to-end delays (hence the name). However, unlike similar types of networks (such as mobile ad-hoc networks), mobility in DTNs is used as an opportunity to deliver data indirectly, in an opportunistic fashion. For example, a node *A* may want to send a message to a node *B*, but will never be in contact with it. However, if both nodes encounter a node *C*, they can use it as an intermediary node in a virtual path. Thus, node *A* would send the message to node *C*, which would carry it until it encounters *B*, and then deliver it. This paradigm, entitled store-carry-and-forward (Conti, Giordano, May, & Passarella, 2010a; Pelusi, Passarella, & Conti, 2006), is the basis of DTNs. There can be any number of intermediary nodes between a message's source and its intended destination.

DTNs are suitable for a large range of applications, including disaster scenarios (Martín-Campillo, Crowcroft, Yoneki, & Martí, 2013; Quispe & Galan, 2014; Torres et al., 2012;), intelligent transportation systems (Gutiérrez-Reina, Toral, Johnson, & Barrero, 2012; Shin, Kim, & Kim, 2012), pervasive healthcare (Conti & Kumar, 2010b), smart cities (Le, Scholten, & Havinga, 2012), advertising (Haddadi, Hui, Henderson, & Brown, 2011), sensor networks (Bharathidasan, An, & Ponduru, 2002), crowd management (EWIDS), context-aware platforms (Dobre, Manea, & Cristea, 2011), wildlife tracking (Juang et al., 2002), internet access in limited conditions (Pentland, Fletcher, & Hasson, 2004), or even interplanetary internet (Akyildiz, Akan, Chen, Fang, & Su, 2003).

The main problem facing data routing and dissemination in DTNs is how to select the next suitable node in a virtual path. The main problem facing data routing and dissemination in DTNs is how to select the next suitable node in a virtual path. From an expert system point of view, nodes in DTN should learn from the

* Corresponding author. Tel.: +34 954 48 12 93.
E-mail addresses: dgutierrezreina@us.es (D. G Reina), radu.ciobanu@cti.pub.ro (R.I. Ciobanu), toral@esi.us.es, storal@us.es (S.L. Toral), ciprian.dobre@cs.pub.ro (C. Dobre).

http://dx.doi.org/10.1016/j.eswa.2016.03.038
0957-4174/

environment in order to make the best possible forwarding decision. The selected data dissemination technique determines the nodes' forwarding decisions. The data dissemination algorithm is fed by the network parameters that can be locally measured to provide as an output the forwarding decisions in every encounter between two nodes. The main challenge is that nodes can only have access to limited information based on their encounters such as social information, friendship, similarity metrics between nodes, topological data like centrality, and so on, to make such forwarding decisions and these parameters can be inter-correlated. Furthermore, the environment is very complex and dynamic due to nodes' mobility and the limitation in nodes' storage of information. Consequently, the system of nodes should adapt quickly and efficiently to the very changeable network conditions.

A basic dissemination algorithm would be to simply spread a message to all encountered nodes that do not have it, in an epidemic fashion (Vahdat & Becker, 2000). However, although epidemic dissemination ensures the lowest possible delivery latency, it may lead to network congestion, since nodes would have to exchange a large amount of data at every contact (especially if the number of nodes and messages generated in the network is high). Moreover, nodes in DTNs are generally mobile devices with limited data memories, so they cannot carry all the messages, leading to buffer overflows and loss of data. On the other hand, selecting a wrong node for relaying data can lead to high delivery latencies and to lower hit rates (i.e. less messages will actually reach their destinations). In addition, the type of network in which a dissemination algorithm is employed in has a great deal of significance in the desired behavior. For example, DTN algorithms employed in disaster scenarios require data to reach the destination as quickly as possible with limited resources, while algorithms used for wildlife tracking need to ensure that all collected data eventually reaches the destination, regardless of the delay.

The goal of this paper is to analyze the various DTN parameters used for dissemination (similarity, centrality, friendship, social strength and trust), and to see how they each affect various dissemination metrics, such as hit rate, latency, or delivery cost. In order to achieve this, we formulate data dissemination in DTNs as a multi-objective optimization problem, and use a genetic algorithm to obtain the Pareto front for the three metrics when varying the importance of the dissemination algorithm parameters. This will allow us to quantify the impact of each parameter on the output of a certain type of DTN, which would in turn permit us to correctly configure a dissemination algorithm based on the network conditions. Furthermore, we provide a selection mechanism based on decision tree for selecting a subset of the optimized solutions that satisfy a given target performance. Consequently, the main contributions of this paper are twofold

- First, we model the data dissemination problem in DTNs as a multi-objective optimization problem that can be solved by an artificial intelligence algorithm like a multi-objective genetic algorithm.
- Second, we combine the multi-objective model and a decision tree based mechanism to obtain the importance and relationship of relevant parameters normally used in the existing data dissemination algorithms. These findings have important implications for the design of new data dissemination algorithms in DTNs.

The rest of the paper is organized as follows. Section 2 presents related work in the field of DTNs and genetic-based dissemination algorithms. Section 3 formulates the data dissemination problem in DTNs as a multi-objective optimization problem, while Section 4 presents the methodology we used and delves deeper into implementation details. Finally, Section 5 contains a thorough analysis of the results obtained, and Section 6 presents our conclusions.

2. Related work

DTN research has been gaining steam in recent years, due to the great increase in the number of existing mobile devices and their ubiquitousness. A thorough analysis of DTNs and opportunistic networks was done by Conti et al. (2010a), where functions such as message forwarding, security, data dissemination and mobility models are analyzed. Several dissemination algorithms are also reviewed, among them being BUBBLE Rap (Hui, Crowcroft, & Yoneki, 2008), PROPICMAN (Nguyen, Giordano, & Puiatti, 2007) or HIBOp (Boldrini, Conti, Jacopini, & Passarella, 2007). A taxonomy for data dissemination algorithms (Ciobanu & Dobre, 2011) splits them based on network infrastructure, node properties, content characteristics, and social awareness. The design of new data dissemination algorithms for DTNs has been a very active research topic in the last decade due to the fact that many decision variables can be considered for forwarding decisions (Sobin, Raychoudhury, Marfia, & Singla, 2016). Another important reason is that there is not a mathematical model in the literature to analyze data dissemination in DTN from theoretical point of view. Consequently many heuristics have been proposed to find quasi-optimal solutions.

More recently, genetic algorithms have also been employed for optimizing many features of DTNs (Gutiérrez-Reina et al., 2016). One such solution, which performs data dissemination in vehicular-based DTNs, was proposed by Bitaghsir and Hendessi (2011). A node decides whether to forward a message to another node it has a contact with by assigning weights to four node properties, and comparing the weighted sum with a threshold. The four properties are the encountered node's speed, its direction, the distance between the node and the message destination, and the probability of network disconnectivity (which is computed from a message's hop count). Based on the four weights (one for each of the above parameters), a first generation of chromosomes is randomly created, each being composed of five genomes: the four weights and the threshold (all values between 0 and 1). For the selection phase, a fitness function that takes into account a message's delivery latency is used. Then, one-point crossover is employed for the reproduction phase, while the termination condition is achieved after four generations. While the solution we propose here uses genetic algorithms, it is more complex, since we use a multi-objective genetic algorithm to obtain a Pareto front and we also consider more parameters when taking the forwarding decisions, such as social, trust and similarity parameters, among others. We are not only interested in the effect of the various DTN parameters on the delivery latency, but also on hit rate and delivery cost, in order to be able to correctly configure our algorithm for a multitude of different types of DTNs. Consequently, we obtain a set of possible solutions that the decision makers or the protocol designers can adjust depending on the target application scenario. As far as the authors know, this is the first work that addresses the data dissemination problem in DTNs as a multi-objective optimization problem.

Another data dissemination technique for DTNs that uses genetic algorithms is GAER (Dhurandher, Sharma, Woungang, Gupta, & Garg, 2014). In this algorithm, a node that wants to relay a message (either created by itself or transported on behalf of another node) applies a genetic algorithm for all the nodes it is currently in contact with, in order to decide to which node the data will be relayed. The initial chromosomes are generated randomly, and they represent an array containing a node's home, office and leisure locations (i.e. its home locations), as well as the home locations of the neighboring nodes. A fitness function used for determining the capability of a node to be the next hop is applied, as well as the other genetic operations, i.e. selection and crossover, until a limited number of generations are computed. The fitness function is