Computer Communications 81 (2016) 12–30

Contents lists available at ScienceDirect

Computer Communications

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

Design and evaluation of a cognitive approach for disseminating semantic knowledge and content in opportunistic networks

Matteo Mordacchini, Lorenzo Valerio*, Marco Conti, Andrea Passarella

Institute for Informatics and Telematics, National Research Council, Via G. Moruzzi 1, Pisa 56214, Italy

ARTICLE INFO

Article history:
Received 29 July 2015
Revised 21 September 2015
Accepted 22 September 2015
Available online 29 December 2015

Keywords:
Opportunistic networks
Cognitive heuristics
Associative semantic network
Data dissemination
Knowledge dissemination

ABSTRACT

In cyber-physical convergence scenarios information flows seamlessly between the physical and the cyber worlds. Here, users' mobile devices represent a natural bridge through which users process acquired information and perform actions. The sheer amount of data available in this context calls for novel, autonomous and lightweight data-filtering solutions, where only relevant information is finally presented to users. Moreover, in many real-world scenarios data is not categorised in predefined topics, but it is generally accompanied by semantic descriptions possibly describing users' interests. In these complex conditions, user devices should autonomously become aware not only of the existence of data in the network, but also of their semantic descriptions and correlations between them. To tackle these issues, we present a set of algorithms for knowledge and data dissemination in opportunistic networks, based on simple and very effective models (called cognitive heuristics) coming from cognitive sciences. We show how to exploit them to disseminate both semantic data and the corresponding data items. We provide a thorough performance analysis, under various different conditions comparing our results against non-cognitive solutions. Simulation results demonstrate the superior performance of our solution towards a more effective semantic knowledge acquisition and representation, and a more tailored content acquisition.

1. Introduction

The physical world is becoming more and more saturated by the presence of a vast number of mobile devices. These devices are able to sense data from the physical environment and autonomously elaborate and exchange information among themselves. This behaviour is leading to a constant and increasing flow of information from the physical world to the cyber one and vice-versa. In this context, data coming from the physical world impacts on the decisions taken by devices acting in the cyber world, whereas the information spread in the cyber world can, in turn, influence the actions taken in the physical world. This complex information scenario is known as the *Cyber–Physical World* (CPW) convergence scenario [1–6].

The users' devices acting in this scenario play a key role, since they represent the "door" to the cyber world through which their human owners can access the massive amount of information available in the cyber world. Indeed, in the CPW convergence scenario users' personal devices can be regarded as the *proxies* of their users in the cyber world. They are most of the time together with their users in the physical world, and can thus, for example, gather context information about their behaviour and the physical places they visit. On the other hand, they are probably the most typical way through which users access information in the cyber world. Therefore, they can be usefully instructed to autonomously act in the cyber world (e.g., by proactively filtering or fetching information) on behalf of their users, by exploiting context information about their behaviour in the physical world. To this end, because the cyber world is typically full of data of all kinds that could possibly be accessed, users' devices should automatically understand which data is important for their users at a given point in time, avoiding to overflow them with useless information.

In this context, the opportunistic networking paradigm plays a relevant role by supporting direct communication between mobile devices. In an opportunistic network, direct, physical contacts between nodes are opportunistically exploited to recognise and disseminate relevant information toward potentially interested nodes, without the need of centralised infrastructures or precomputed paths from source to destination [7–10]. Beyond the problem of data dissemination, it is worth to mention that the other main research issues in opportunistic networks focus on the development

* Corresponding author. Tel.: +39 50 315 3059.
E-mail addresses: m.mordacchini@iit.cnr.it (M. Mordacchini), lorenzo.valerio@iit.cnr.it, lorenzo.valerio@gmail.com (L. Valerio), m.conti@iit.cnr.it (M. Conti), a.passarella@iit.cnr.it (A. Passarella).

http://dx.doi.org/10.1016/j.comcom.2015.09.027
0140-3664/

of analytical models of data delivery performance [11–13], routing approaches that consider nodes' aggregation and privacy [14,15], forwarding schemes for hybrid networks (opportunistic networks combined with the infrastructure [16]), real world implementations [17], applications to Vehicular Networks [18,19]. Recently, opportunistic networks have been proposed as one of the possible key components of future mobile networks (e.g., in the 5G domain), as they are able to complement wireless infrastructures such as cellular and WiFi networks, by enabling direct dissemination of data among users nearby, thus contributing to offload data from the cellular network [20,21].

Generally, in analysing data dissemination in opportunistic networks it is assumed that the users' interest is rather static and quite simple to describe [22]. In typical data dissemination approaches, users are supposed to be interested in predefined content categories (e.g., sports, movies, etc.) and therefore their devices collect all the contents related to those categories. An aspect that has not been often taken into consideration in the literature on opportunistic networks is that contents are also equipped with rich semantic descriptions (i.e. associated concepts and/or tags). This could be the case, for instance, of tagged photos on Flickr and Instagram, or messages annotated with "hashtags" in Twitter and Facebook. Often, users' interests toward contents are driven by the content descriptions, which triggers interest in contents themselves: data items[1] are accessed by users because their semantic description contains information that is relevant to the user at that moment. Therefore, our key idea in this paper is to optimise in an intertwined way the dissemination of both *semantic* data associated to contents, and of *contents* themselves. Specifically, in the proposed approach semantic data is disseminated among users' devices, based on the current interests of the users. By receiving semantic data, users' devices also know what contents (associated to that semantic data) are available in the network, and fetch them, if appropriate. Therefore, in our approach users' (dynamic) interests drive the dissemination of semantic data, which drives the dissemination of content. As we explain in Section 3, this mechanism is quite similar to the way users access information in the physical world.

Specifically, in this paper we apply concepts coming from the cognitive science field, and design a system whereby users' devices autonomously become aware of the structure of the semantic information describing the available content in the environment, and disseminate contents based on this knowledge. In particular, we show how devices can exchange information in a way that resembles how conversations between humans enable spreading of ideas (i.e. semantic information), which generates interests for specific types of content, and ultimately determine content that people access. To this end, we consider that, acting on behalf of their users inside the cyber world, mobile devices are exposed to problems similar to the ones faced by the human brain when dealing with the information and content selection tasks. Cognitive scientists produced, during years, many functional models and descriptions of these mental schemes. These functional models, called *cognitive heuristics*, differently from other biological models as artificial neural networks, do not aim at reproducing the physiology of the brain's processes, but model their functionality. By taking advantage of these descriptions, previous works [23–25] have shown how rules and procedures used by the human brain, when assessing the relevance of information (in face of time and resource restrictions), can be exploited to design adaptive, low resource-demanding, yet very effective, algorithms for data dissemination in opportunistic networks but none – to the best of our knowledge – has explored how to apply these models to optimise the joint dissemination of semantic information and associated contents.

In order to take advantage of these models, we have to face the problem of how to represent semantic information in devices' memory, how semantic information is retrieved and exchanged upon contacts between nodes in physical proximity, and how content is finally selected for dissemination, based on the semantic data exchange that has been carried on. For each node, the internal memory representation of semantic concepts is inspired by the associative network models (AN) [26,27] of human memory coming the from cognitive psychology field. In AN models, semantic concepts are represented by nodes that are interconnected by paths that vary in strength, reflecting the degree of association between each pair of concepts. In our proposal, each mobile node builds a local semantic representation of its own contents through a semantic directed weighted graph, where vertices represent the semantic concepts associated to data items, and the edges represent the semantic relationships between concepts, both learnt from the environment and derived from the node's own contents. Moreover, since memory is a limited resource at each node, cognitive models of how the least relevant information can be dropped from memory [28,29] are exploited. When nodes come in physical proximity (are in contact), exchange of semantic information happens as follows. Communication takes place only between nodes having some common interest, i.e. only if there is an initial set of semantic concepts shared by both nodes. Concepts to be exchanged are selected by navigating each semantic network, according to an edge ranking algorithm derived from the *fluency* cognitive heuristic (FH) [30,31]. Finally, we show how mobile nodes, after having enriched their semantic graphs with new concepts taken from other encountered nodes, select which contents, locally available at one of the nodes, to exchange between them, giving precedence to contents whose semantic information maximally overlaps with semantic concepts just exchanged between nodes. In cognitive terms this refers to the *tallying heuristic* (TH) [32], another cognitive decision strategy used by human brain. All these cognitive processes are described in more details in Section 3.

The rest of the paper is organised as follows. In Section 2 we present state-of-the-art data dissemination approaches for opportunistic networks. In Section 3 we show at a high level how each cognitive model relates to our solution. In Section 4 we present the entire approach in full detail. In Section 5 a thorough performance analysis of our cognitive based system is provided and, finally, Section 6 concludes the paper.

2. Related data dissemination approaches

The problem of data dissemination in opportunistic networks have been addressed by many works during years. The first attempt was in the context of the PodNet Project [33]. The authors proposed a solution where nodes cooperatively exchange data items in order to retrieve all those contents they are interested in. Precisely, contents are organised in predetermined channels of interest to which nodes are subscribed. In order to favour the data dissemination, upon encounter, nodes load in a *public cache* items they are not directly interested in. Items to be maintained in the public cache are chosen depending on different history-based strategies that consider the past received requests, interpreted by nodes as a popularity index of the channel of interest. These strategies could be effective when users mobility is homogeneous and contents can easily traverse the network. However, this approach suffers in scenarios where nodes tend to group in communities and their movements are heterogeneous.

Advances in data dissemination solutions leave the content-centric approach adopted in PodNet in favour of more user-centric solutions. Precisely, these solutions define more elaborate

[1] In the paper we use the terms content and data item interchangeably.