Ad Hoc Networks 47 (2016) 53–81

Contents lists available at ScienceDirect

Ad Hoc Networks

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

A survey on clustering techniques for cooperative wireless networks

Victor Sucasas[a,b,*], Ayman Radwan[a], Hugo Marques[a,b], Jonathan Rodriguez[a], Seiamak Vahid[b], Rahim Tafazolli[b]

[a] Instituto de Telecomunicações, Aveiro, Portugal
[b] Institute for Communication Systems, University of Surrey, UK

ARTICLE INFO

Article history:
Received 9 July 2015
Revised 22 April 2016
Accepted 23 April 2016
Available online 3 May 2016

Keywords:
Clustering
Ad hoc networks
Cooperative networks

ABSTRACT

Clustering became relevant in the past as a solution for the scalability problems of ad hoc networking, but, the unsuccessful application of ad hoc solutions to real scenarios, such as the projects SURAN and PRNet, decreased the interest of research community on ad hoc communications, and subsequently, on clustering algorithms. Recently, however, clustering techniques have gained renewed interest due to the emergence of cooperative communications for cellular networking. Clustering is envisaged, in this scenario, as a technique to team up nodes to support efficient data aggregation for energy saving, scalability and privacy among other benefits. Moreover, research on 5G networks also envisages a connected society, where everything and everyone will be connected under the umbrella of Internet of Everything (IoE). This novel communication paradigm has fostered new research on clustering, which has yielded novel and more advanced algorithms and applications. This article surveys the State-of-the-Art in clustering techniques and provides detailed descriptions of the basics of clustering and the latest novel ideas. Open issues, technical challenges and directions for future research are also outlined.

1. Introduction

Clustering, in telecommunication systems, appeared in the framework of ad hoc networking, which was proposed decades ago as a promising solution for infrastructure deprived networks, such as military battlefield communications, emergency communication black-outs or to deploy fit for purpose wireless networks [1]. The idea of setting a network, on-demand, without implementing access points or base stations, was very appealing from both the utility and economic perspective. In ad hoc networking, the nodes communicate among themselves by transmitting and relaying the packets from source to destination, without a common hub that routes the packets or organizes the communication flow. This paradigm, however, requires complex routing mechanisms to optimize the route path selection and ensure a satisfactory packet delivery ratio. This was indeed the bottleneck of ad hoc networking, since the unpredictable nature of this scenario, where nodes show up and vanish continuously, requires a continuous update of the routing information. This information can grow significantly in big networks, due to the countless route paths that routing algorithms must evaluate. Clustering was suggested in this framework to reduce the routing complexity and provide scalability. The strategy of clustering is teaming up nodes in virtual groups, with one leader per group, such that routing techniques consider these groups as sole entities, hence reducing the number of network sources, destinations and possible paths considerably, hence increasing its stability.

The research interest in clustering techniques went side by side with the concern of ad hoc networking, very prolific during two decades, producing research efforts in a variety of scenarios and technologies such as mobile wireless networks and sensor networks. This interest however has been gradually declining due to the absence of implementation of ad hoc solutions in real scenarios. With the sole exception of wireless sensor networks, ad hoc networking has been considered a liability problem, and the main projects on ad hoc solutions for military and emergency situations have failed [2,3]. Clustering, originally designed for ad hoc networking, lost some relevance for research community. However, the emergence of more advanced wireless devices such as smartphones, wearable technology and vehicular networks, provided with higher processing capabilities and multi-standard wireless interfaces, has catered a new scenario for clustering in the framework of cellular networks and cooperative communications. In this scenario, cooperation provides many benefits and only one requirement. The benefits include energy efficiency, scalability, efficient location services, social networking and privacy preserving among others. The requirement is an efficient technique to establish cooperative groups of nodes, i.e. clusters.

* Corresponding author at: Instituto de Telecomunicações, Aveiro, Portugal.
E-mail address: vsucasas@av.it.pt (V. Sucasas).

http://dx.doi.org/10.1016/j.adhoc.2016.04.008

Energy saving can be accomplished by using different wireless standards for the short range and long range links, i.e. a low energy demanding wireless interface, such as Bluetooth, can be used in short range to aggregate traffic, whereas a higher range and higher energy demanding link will be used to relay this traffic. Clustering is in this scenario envisaged to provide a stable cooperative framework between nodes [4]. Multi-standard wireless transmission is not the only way to achieve energy efficiency, nodes can also transmit information, in short range links with high transmission bit rates, to the relay nodes with better channel conditions, hence with the capability to transmit at high order modulation for longer periods, which effectively reduces the transmission time [5]. This strategy has been proposed in different scenarios, including wireless sensor networks [6] or mobile wireless networks [7]. Moreover, efficient location services [8] and mobile social networking [9] for MANETs can be supported by cooperative clustering.

Vehicular networking also considers the approach of teaming up nodes for scalability [10] and privacy reasons [11]. Traffic monitoring applications require a continuous transmission of information updates from vehicles to a backend server, hence this information goes through the base station. According to current analysis of the state of the art, this continuous transmission of information can form a bottleneck in the LTE uplink connection. Data aggregation has been proposed in this scenario to reduce the number of uplink connections [10]. In short range, vehicles can transmit and aggregate traffic using the Dedicated Short Range Connectivity (DSRC, 802.11p) wireless interface, then the relay nodes transmit the aggregated traffic over the LTE uplink. Regarding privacy preserving for vehicular networks, it can also be provided by aggregating nodes into clusters, since data aggregation hides the identities of the nodes forming the cluster (only the identity of the leader of the cluster is revealed) [11]. Internet connectivity is also envisaged as a desirable commodity for VANETs, where again QoS in terms of jitter, delay, packet delivery ratio or connection duration are the most relevant metrics. These requirements have inspired several studies on the boundaries of the achievable QoS for VANETs [12]. Clustering, in this framework, can increase the QoS by providing more connection reliability and increasing connectivity in low coverage scenarios. The Section 15 of this survey provides some examples of data aggregation algorithms in novel scenarios such as IoT, Smart City and VANETs.

Clustering is also envisaged as a tool to provide connectivity in hybrid wireless networks [13]. In a hybrid wireless network, nodes self-organize into a hierarchical (clustered) ad hoc network, and uplink connections are provided by scattered nodes to ensure connectivity in case of network partitions and also to provide internet access. In this scenario, clustering can be used to select the nodes that establish the uplink connections, i.e. the clusterheads, that connect to the base station and provide a gateway for the cluster-members.

Fig. 1 depicts a number of examples of the clustering applications in these emerging scenarios. Smartphones, wearable devices and vehicular technology—cellular networks in general—apply clustering to coordinate the nodes into groups to support efficient cooperative communication strategies. These new scenarios have however yielded new requirements for clustering, that must adopt a multi-objective evaluation strategy for cluster formation, where nodes hardware capabilities, mobility patterns, channel availability, among others, must be taken into account to form clusters. Previously proposed algorithms were mainly based on the consideration of one only parameter such as the node identity, number of neighboring nodes, battery level, speed, etc. In the new scenarios, the network is populated with different technologies (vehicular networks, mobile networks, body area networks, sensor networks, etc.), a variety of hardware capabilities, and the nodes are highly mobile. These technologies are not isolated from each other and it is common to have different devices sharing the same scenario, hence novel clustering techniques must consider a heterogeneous network, where clusters are formed by nodes with different characteristics.

This survey offers a retrospective walkthrough on clustering techniques, with detailed insights on the origins of clustering and the most relevant research efforts during the previous decades. We detail in this manuscript the main merits of previous proposals but also the main drawbacks, and the limitations of these algorithms to be applied in novel areas of wireless networking. The survey also addresses how the new scenarios of wireless networking have fostered a renewed research on clustering, and describes the most recent approaches. At the end, the survey also concludes with the learned lessons that can be drawn from surveying the state of the art in clustering.

2. Organization of the survey

This manuscript is organized as follows: Section 2 defines what is clustering, and provides a thorough explanation of its functionality and main objectives; Section 3 describes the clustering benefits, costs and design considerations; Section 4 describes the most adopted metrics for clustering performance evaluation; Section 5 provides a thorough walkthrough of the origins of clustering; From Sections 6 to 14 the manuscript lists the different clustering categories and describes the most novel algorithms falling in each of these categories. The survey also provides summary tables at the end of each section where readers can find a short description of the algorithms detailed in the text; Section 15 provides a short summary of data aggregation in clustered topologies; Section 16 provides the lessons learned by the authors of this manuscript; Section 17 describes the newest trends and future research directions; finally, Section 18 concludes this survey.

Fig. 2 provides a schematic chart of the main approaches in the state-of-the-art of clustering described along this survey. This figure depicts the main fields, respective sub-fields, and the newest trends of clustering, and includes some references falling in these research areas. In this schematic chart, the tagged squares represent the main clustering categories, from which the respective subfields are extended representing the evolution of clustering techniques along the last two decades. Fields included in circles represent future trends and new topics.

3. What is clustering and how does it work?

Clustering techniques are diverse, but there is a common trend followed by the majority of research efforts that was given by the first publications on clustering. This common approach is based on converting a flat network into a clustered network by assigning roles: clusterhead (leader of the cluster); cluster-member (node belonging to the cluster); and gateway (a cluster-member that belongs to one cluster but connects two clusterheads). The common approach is to select the clusterheads first according to a predefined metric, then the rest of the nodes join one of the surrounding clusterheads and finally the cluster-members that interconnect clusters become gateways. Fig. 3 depicts this procedure.

The clustering formation process requires the transmission of control information. Nodes transmit information such as node identity, list of neighbors, mobility, battery level, etc. This information is required for the clusterhead selection and cluster formation process. The node identity and the list of neighbors is a common control information in all clustering algorithms; other metrics such as mobility or energy, however, are characteristics of some algorithms that target very specific scenarios, such as mobile networks