Design Methodology for Self-organized Mobile Networks Based

John Petearson Anzola, Sandro Javier Bolaños and Giovanny Mauricio Tarazona Bermúdez

Universidad Distrital Francisco José de Caldas, Bogotá D.C., Colombia

Abstract — The methodology proposed in this article enables a systematic design of routing algorithms based on schemes of biclustering, which allows you to respond with timely techniques, clustering heuristics proposed by a researcher, and a focused approach to routing in the choice of clusterhead nodes. This process uses heuristics aimed at improving the different costs in communication surface groups called biclusters. This methodology globally enables a variety of techniques and heuristics of clustering that have been addressed in routing algorithms, but we have not explored all possible alternatives and their different assessments. Therefore, the methodology oriented design research of routing algorithms based on biclustering schemes will allow new concepts of evolutionary routing along with the ability to adapt the topological changes that occur in self-organized data networks.

Keywords — Self-Organizing Networks, Routing Methodology, Biclustering Algorithms.

I. INTRODUCTION

One of the everyday questions in self-organized data networks is this: are the groupings needed? The answer is evident in the increase in overall performance that presents cluster topologies in comparison with non-clustered topologies. When using a routing protocol, mobile nodes are organized into cluster structures that facilitate access and global management of the network. With a cluster structure, the network can answer topological changes caused by the mobility of the nodes [1]–[6]. Below are several advantages in this type of structures:

1. A structure of groups facilitates the reuse of resources in space to increase the capabilities of the system. Non-overlapping, unconnected multi-cluster arrangements inside can be deployed with the same frequency or channel of communications to reduce collisions of data transmission in the network [7]–[9].

2. Structure of groups routing schemes use a set of nodes called clusterheads. These function as gateways that distribute the traffic to the inside of each group and allow interoperability with neighboring groups, so dissemination of data is decentralized [10]–[12].

3. Structure of groups make a network that is smaller and more stable, according to the employed group heuristics [13]–[15].

4. For events of admission of a node to a group, nodes residing in the group that have just entered and the group from nodes, update information routing. In this way, for groups not involved, nodes do not observe these changes, which reduces the burden of traffic generated in the network.

Another technique in a different context that allows for grouping is Microarrays, which represent a new way of measuring levels of gene expression under different biological conditions in multiple data. These data are successfully analyzed by methods of biclustering, which removes a number of genes and conditions that show a similar behavior [16]–[18].

Biclustering techniques are conventional clustering techniques for simultaneous genes and conditions grouping characteristics, as well as by relations of overlap between groups. In simultaneous clustering of genes and conditions, the groups found by a biclustering method are called biclusters and refer to the genes that act similarly under a given set of conditions. This observation of the biological condition in a group of genes behavior characterized relationships between genes to a number of conditions and its decoupling under other conditions. In the overlap feature, you can see genes in more than one bicluster, which is interesting, since in biological observations, a gene can have more than one role and relationship associated with different sets of genes and conditions.

In this article, we propose a routing methodology based on an algorithm of biclustering for self-organized data networks, under an evolutionary approach that allows biclusters to be highly correlated. The evaluation of the methodology proposed by a researcher is a function of the conception of the heuristics of grouping under the conditions of the types of data and services that are to be used, which for the case study intended to stage the Internet of Things (IoT). Throughout the article, there is an isomorphism between the biclustering algorithms employed in the analysis of gene expression; similarly, it can assimilate in the clustering of nodes in mobile networks ad hoc, referred to as self-organized data networks, showing how the proposed methodology can build a routing protocol.

II. PROBLEMS IN SELF-ORGANIZED DATA

One of the problems treated in self-organized data networks relates to obtaining location, tracking and metrics deployment of nodes based on the spatial coverage, highlighting the example of fire detection, where you can ask as a particular case:

- With that can quality of service (QoS) nodes in the network monitor a specific area?
- Can the network draw out info from the specific location of the fire and determine the propagation time?
- How should nodes be distributed to monitor the largest possible coverage area?
- Can a structure in groups reply in the shortest time compared with individual collaborative work from all nodes?

Each of the above questions can reveal weaknesses in physical coverage areas and suggest a scheme of movement, deployment and reconfiguration of position, the motion of the nodes can improve the quality of the information provided. For the above example, not only the physical location involved in improving the response time and the quality of information, there are other factors that affect communication between nodes, which may arise directly or indirectly:
• How to access the communications channel, if all nodes to transmit at the same time?
• What is the proper path to send the information to minimize data loss?

These questions cover issues treated at the physical layer, link and network, whose model in layers and interaction carried out among them respond to most of the problems encountered in data networks, being a communication protocol which brings the interactions between layers to mitigate these problems.

In terms of data transmission according to the routing, the path of greater power and range of coverage between two end-to-end terminals facilitates communication and efficiency, to minimize power consumption and maximize the throughput. These features of coverage in terms of space and range of data transmission, have multiple problems that have been addressed by grouping schemes [19]–[21].

On the other hand, accessibility in self-organized data networks is limited by interference from simultaneous transmissions to share a unique medium, which affected the performance by the constant struggle of the channel, without adding effects and delays caused by the topological changes related to mobility. These problems were discussed with clustering schemes that enhance the results of the scalability of routes depending on the density of the network, since they minimize the effects of on long-haul routes in contrast to routing without clusters schemes [21], [22].

Another problem in self-organized data networks is due to planning and booking of resources. One must consider the medium access control (MAC) protocols, which are responsible for coordinating access to the active nodes to a wireless environment that is prone to errors, not to mention that the search and discovery of routes with high frequency and coverage presents the hidden terminal problem [23], [24]. Therefore, these actions may be considered problems of planning policies to access the environment and provision of resources. In the link layer, planning is treated by the prioritization of packages and services, which is the case of the mechanism enhanced distributed channel access (EDCA) [25], [26]. In the network layer protocols that are based on clustering, routing schemes allow for addressing problems of planning in large networks in order to achieve better quality of service.

In this section, the problems in self-organized networks of data have been synthesized and classified into four categories: mobility, coverage, planning, and topology control, as shown in Fig. 1. The problems of classification shown in Fig. 1 are common pattern routings [53], which become a transverse solution to the problems presented. At the same time, routing protocols are classified according to the way of discovery and selection of routes in:

• Proactive: In this type of protocols, periodically a server node issues a “Hello” packet to the network to investigate that nodes are born, live, and die in the network, through the construction of a route towards them, with the assumption that at some point the alternative routes will be needed and used. The “Hello” package allows you to update the routing tables either by changing the position or death of a node in the network.

• Reactive: The reactive part of a protocol occurs when a node wants to find a path to a node destination through flood processes; this process is called route discovery. Once the proper path, it remains there until the destination node becomes unavailable, usually by a topological change or loss of the trajectory. The occurrence of a change or loss event obligates the route discovery process to start again.

• Hybrid: hybrid protocols combine the processes carried out by approaches to routing proactive and reactive, simultaneously leading routing for the intragroup and intergroup [27], [28].
Self-organized data networks within the source nodes are wireless mobile devices that are added to the network, allowing the growth of “things.” These devices are added and adapt autonomously after an ad hoc configuration, if they have been configured in infrastructure mode. These devices are also candidates to serve as target nodes by the organization’s edge that are present within the network.

The scenario described in Fig. 2, is represented in Fig. 3, through an enterprise architecture that generally refers to the type of work that takes place in an organization, that describes, are not only technological parts of hardware and software, but also users and processes. At this point, the architecture of business models using the technology, users and processes in the full context of an organization and its interaction with the business[37]. The case tried in Fig. 2, the interoperability between wireless networks and the internet of things is described as a model of enterprise architecture in Fig. 3.

Fig. 3. ArchiMate Infrastructure model.

IV. DYNAMIC INFRASTRUCTURE MODEL BEHAVIOR

In order to understand the dynamic behavior of the model Fig. 3, describes the internal processes of communication protocols. Communication from the end user to the target node is an IP communication. Communication from the target to the source node consists of processes of flooding, clustering heuristics, data transfer, exceptions, maintenance and completion of the transfer of data, which are represented from Fig. 5 to Fig. 9, using sequence diagrams in unified modeling language (UML). In Fig. 4, each of the interactions and processes is shown in the time, detailing the flow of different packages that are carried in a self-organized data network or ad hoc mobile network, with clarifications to the level of messages from existing objects.

Fig. 4. General processes in a mobile wireless network.

Model Fig. 3, is based on the premise that “the end-user has internet connectivity”. Once connectivity is guaranteed, we proceed to implement the general application design model client/server, following the establishment, data transfer, and completion of the communication processes. This model is known as a three-way handshake, creating a virtual logical circuit from end to end between the end user and the target node. On the model of Fig. 3, the server node is the target that centralizes the following services represented by sequence diagrams:

**Flood process:** Consists of two packages, the package Request Route (RREQ), which is intended to build a path forward from the target node to the source node. Once the RREQ packet arrives at the source node, this returns a Route Reply (RREP) response packet to the node target, building a route back. Once the RREP packet arrives at the target node, it generates an array of routing with genes that are the data collected in the course of back and forth, becoming an information input into the process of biclustering.

**Heuristic grouping process:** Once the target node has received the package RREP containing genes or data collected, in the process of flooding an array of routing is built in conjunction with the conditions of the service that is required in the network. These conditions determine the end user making the request for service to the target nodes, which are distributed at the edges of the network and evaluate the proximity of the node source consulted by the end user.

Once the heuristics of grouping is complete, proceed to flood the network with package Map REQuest (MREQ). This flood is directed to locate nodes clusterhead and establish the routes and nodes inside each bicluster. Thus, the nodes that make up each bicluster are associated and related to the genes and conditions of the network, whose distribution is coordinated with the clusterhead in routing, supported in the heuristics that respond to the solution of one or several of the problems recorded in Fig. 1.

Once the heuristics of grouping is complete, proceed to flood the network with package Map REQuest (MREQ). This flood is directed to locate nodes clusterhead and establish the routes and nodes inside each bicluster. Thus, the nodes that make up each bicluster are associated and related to the genes and conditions of the network, whose distribution is coordinated with the clusterhead in routing, supported in the heuristics that respond to the solution of one or several of the problems recorded in Fig. 1.

As heuristics, in the implementation of the biclustering algorithm you can select a two nodes clusterhead, in order to provide redundancy in each bicluster, given the case that if one of them fails, you have routing information backed up on a second node that would become clusterhead. This second node subsequently must run within the bicluster a clusterhead of backup node selection process.

**Data transfer process:** This process is responsible for starting the bi-directional communication between the target and source nodes. Data transfer is initiated by the target node since it chooses the path to the node source through the issuance of a packet Data REQuest (DREQ). Once the service request is serviced by the source node to
receive the package DREQ, it emits a packet Data Reply (DREP) with the response of the service requested by the target node. The target through a process of interoperability node sends the end user terminal DREP package with the data required by the user. This process can occur several exceptions, which is the loss of packets, routes, and events for the birth and death of a node in the network.

**Maintenance process:** This process is conducted by clusterhead and target nodes, consists in maintenance of routes and the collection of information of genes within each bicluster, programming a local and periodic flooding of packets Hello REQuest (HREQ) shipments. This package allows you to get the information of genes within the bicluster to monitor routes and active nodes, as well as events of admission, birth, and death of a node. In the event of a node to the bicluster entry, the node receives HREQ package and generates a package Hello Reply (HREP) with genes, informed its current state within the bicluster to the target node. This process makes it possible to act like a lasso feedback between nodes clusterhead, source, and target, allowing the target node to execute the algorithm of biclustering generated a lapse feedback, which subsequently allows the node target adjustments or decisions.

This maintenance process allows foreseeing events within the bicluster feedback with HREQ and HREP packages. This feedback allows you to monitor states of links, the different routes and different metrics that can extract the package HREQ. When presenting an event that leads to the network on the brink of chaos, part of the overall structure is lost, either by overflow of data, a drop of energy in several nodes, abrupt movements of nodes or any other event that generates chaos, in any case, the network should be able to adapt to the new changes or gradual transformations seeking a balance and a reorganization. This maintenance process allows the network to evolve.

**Finalization process data:** In this process the source node sends a packet REQuest End (EREQ) to the target node and the target node responds with a package End Reply (ERE) to the source node by the main routes and alternative routes active, in order to free up resources if these were used. At the same time, the target node sends a DREP packet that tells the end user terminal that the data transfer has been completed.

### V. Biclustering Algorithm

Biclustering is an unsupervised data analysis technique that has been applied in studies of gene expression. This technique is a natural evolution of the clustering [38] and its term was introduced by Mirkin [39], later strengthened by Cheng and Chu-Hsing [40]. This technique comprises sets of genes with a similar genetic profile in all the experimental conditions tested.

The analytical capacity of the biclustering analysis of gene expression is greater than the results obtained by techniques of traditional clustering, clustering simultaneous genes and conditions, and the overlap that may occur.

The groups found by simultaneous clustering of genes and conditions are referred to genes that act in a similar way under a single set of conditions, not necessarily all, which fits with the observed biological behavior. Under these conditions gene groups can work together to attend a particular circumstance, but be uncoupled under others.

Overlap allows handling genes into more than one bicluster simultaneously, since the biological reality of one or several genes may have more than one associated function and work with different sets of genes under different conditions.

Similar to traditional clustering techniques, the possibilities for the calculation of similarity within a bicluster have been identified in four main structures within the sub-arrays groups, as shown in Fig. 10.

The package of the levels of expression of genes with a condition or sample form a vector called the profile of that condition. Gene expression profiles are powerful sources of information that are organized in a matrix whose rows correspond to genes and columns to the conditions. One or more objectives of common analysis are a group of conditions and genes are subsets that convey a meaning of biological [41].

The resulting subsets called bicluster can be computationally interpreted as clusters allowing you to group a set of genes, which are linked to certain conditions of similarity measures simultaneously intragroup and intergroup, whose overlapping provides information of relations between groups, as with traditional clustering techniques they are not able to identify.

![Fig. 10. Different bicluster types.](image)
Then is an isomorphism of biclustering algorithms implemented in problems of gene expression with the problems encountered in self-organized data networks, linking them in the following form:

- Genes are isomorphic to the metrics of input, such as:
  - Number of hops
  - Received Signal Strength Indicator (RSSI)
  - Round-Trip delay Time (RTT)
  - Packet delivery ratio (PDR)
  - Bit Error Ratio (BER)
  - Energy consumption
  - Coordinates GPS

Note: At this point, the researcher can propose metrics or heuristics of input.

- Conditions are related to restrictions of service or network, such as:
  - Latency
  - Throughput
  - Peak Signal-to-Noise Ratio (PSNR)
  - Shortest route
  - Quality of Service (QoS)
  - Quality of Experience (QoE)

Note: At this point, the researcher can propose restrictions of heuristics that will require service.

Genes are obtained by flooding and maintenance processes, the target node collects the RREP and HREP packages to feed the matrix of routing, which subsequently applies a routing algorithm for biclustering restrictions that are needed to manage the network.

The biclustering algorithm compiles a set of genes that form a matrix of attributes of each node, when applying the algorithm proposed by the investigator, to get one or several sets of nodes are called biclusters, which are commonly disjointed edge or border nodes called nodes clusterhead, whose function is the interoperability between clusters. To evolve the biclustering algorithm are overlapping biclusters presenting similar and shared attributes between the biclusters obtained previously. Overlap found in these biclusters feature allows them to respond to given conditions or restrictions routing heuristics that demand services and network traffic.

The methodology proposed in Fig. 12 responds to the problems posed by the complexity of self-organized data networks, ranging from the search for the best route to the optimization of multiple targets, which can work under approaches to evolutionary algorithms specializing in clustering [50], as those recorded in Table 1 and/or under the classical approach of evolutionary techniques [51] recorded in Table 2.

Fig. 12. Design methodology of self-organized networks.
B. Description of the methodology

The characterization of the IoT part of principle than anything else anywhere has connectivity to the internet, based on the above, within the methodology proposed in Fig. 12. The end-user has connectivity to the target node. Self-organization of a network is given by “things”, or wireless mobile nodes, that adaptively can connect to the internet as a principle of functionality in self-organized data networks.

Target nodes are nodes that are organized to the edge or border of the internet and allow interoperability between self-organized networks and the internet. The establishment of communication between the source node and the target node is given by a series of algorithms that the designer of networks should consider until a communication point to point is established. This communication is not done by human beings, but is a communication made by two machines; therefore, communication protocols can be considered an M2M (Machine to Machine) communication.

Consideration 1, algorithm of flood. The initial section of the methodology is composed of a series of floods that are estimated as simple routing algorithms. The methodology consists of two types of available flood: uncontrolled flood and controlled flood.

In the uncontrolled floods, all nodes send packets to their neighbors indefinitely and with more than two neighbors, create a storm of broadcasting.

In controlled floods, there are reliability rules, such is the case of sequence number controlled flooding (SNCF) and reverse path flooding (RPF). In SNCF, nodes are attached to an address and number sequence in a package and are to be transmitted. These are stored in the buffer for each node by attaching your address and your sequence number, so that when you receive a package with the same address and number of the origin node stream, this is rejected. On the other hand, in RPF the nodes send the package forward, if the packet is received by a node this is sent back to the sender node and emitted a series of package forward, if this is received, a package to return to the node which sent him and this in turn emits the same package to the node that sent the first package.

Neighbor discovery protocol (NDP) used in IPv6 is similar to address resolution protocol (ARP) used in IPv4. Both protocols allow a node to determine the direction of link layer (MAC address) of a node which is just entering the network; the issuance of this broadcast package allows you to discover the presence of other nodes on the same channel, determine their addresses MAC and the maintenance of the information of the active nodes on the network connectivity. The main objective of flooding is to collect the largest amount of information; collected information metrics become genes and an increase in the number of genes is likely to become a more selective and adaptive bicluster.

Consideration 2, type of routing. The design of self-organized networks requires a perspective of routing that can be proactive, reactive or hybrid. A proactive routing perspective is characterized by search paths regularly with the assumption that these will be useful in the future, while the reactive perspective only seeks a route when it is necessary, on the other hand, the researcher can give a tint hybrid combined the two perspectives.

Consideration 3, selection problems in self-organized networks. The next part of the design required to select one or more of the problems seen in Fig. 1, which are classified in coverage, mobility and topology planning. Once you choose the problem, it is associated to a set of genes; the conditions requiring a biclustering algorithm that controls action is estimated at routing and why controlled release means you can manage the topology.

Consideration 4, construction of the routing matrix. The construction of the routing matrix is based on the conditions of the network and type of service. At this point, you select the heuristics that allow the creation of biclusters to optimize the routes depending on the problems addressed in the consideration 3. The selected genes become monitoring metrics that monitor the state of resources of the biclusters; the conditions become restrictions on the parameters of quality of service. This consideration allows conceiving with an overview of the design of the routing algorithm.

Consideration 5, type of transport. The type of transport is associated with the service that you want to get from the source Usually within the networks node IP are two types guidance, protocols such as transmission control protocol (TCP) connection-oriented and not as user datagram protocol (UDP) connection-oriented. This methodology aims to experiment on the quick UDP internet connections (QUIC) protocol that was conceived by Google in 2012 and implemented in the
Consideration 6, feedback. In the feedback loop there is a module of packet monitoring that assesses flood packages that monitor the state of the network, with a frequency of monitoring defined by the researcher. Request a route, reservation of resources, evaluation of QoS, birth, life, and death of a node parameter is identified within this module. Feedback is only made by clusterhead nodes and target nodes to avoid the overload of information on the net.

VII. CONCLUSIONS

An approach using hybrid allows you to seize the advantages of proactive approach intragroup, allowing nodes clusterhead to obtain genetic information in some few jumps. On the other hand, maintenance of the bicluster nodes decreases overload of information across the network, to monitor the status of the routes, failure in links, processes of birth, events in life such as admission or exclusion of a node to a group or the death of the same are events that recharged clusterhead nodes.

The focus hybrid to appropriate approach reagent intergroup enables the target nodes limit the total number of nodes clusterhead since the global routing information is maintained between nodes target and clusterhead, reducing the size of the matrices of routing between nodes of the bicluster. Because of the use of hybrid approaches to routing, weaknesses in latency for routes in nodes clusterhead restoration process are introduced in the methodology and win robustness in adaptability and scalability processes.

Regardless of the problems in data networks, self-organizing, clustering structure always improves the overall performance in the network from structures that are not composed of clustering, therefore the proposed routing methodology brings the advantages of using a biclustering algorithm compared to traditional methods, given by clustering techniques simultaneously gene presenting relations under certain network conditions, similarity measures operate intragroup and intergroup, finding overlap that provide information of relationships between groups that fail to identify with traditional clustering techniques.

An advantage of the biclustering algorithms is related to the consistent evolution, a product of a similar behavior in forming operations of bicluster, which through the feedback process treated in consideration 6 of the methodology, algorithm based information from genes and conditions can adapt to the changes that are occurring.

REFERENCES


John Petearson Anzola Teaching program of Electronic Engineering University Fundación Universitaria Los Libertadores, MSc in Information Sciences and Communications University Distrital Francisco José de Caldas, is currently studying the first year of PhD Engineering of the University Distrital Francisco José de Caldas, Bogotá D.C., Colombia. Integral is Research groups GICOECOL and GUIAS.

Sandro Javier Bolaños Doctor of the Pontifical University of Salamanca in Madrid, Spain and PhD Prize awarded by the university; entitled as a systems engineer and magister in teleinformática of the University Francisco Jose de Caldas in Bogota Colombia; a member of the international research group GICOGE researcher in Software Engineering, Software Engineering faculty at undergraduate, postgraduate and expertise in the area of Software Engineering; curricular projects director of undergraduate and graduate, now runs the graduate in Software Engineering and Computer Project at the University Francisco José de Caldas.

Giovanny M. Tarazona Bermudez is a Ph.D. ABD in systems and computer services for internet at the Oviedo University, Asturias, Spain (2012). He acquired a Diploma of advanced studies November 2009 of the Pontifical University of Salamanca. He is an Industrial engineer and Professor of the Francisco José de Caldas District University. Director Research Group on Electronic Commerce Colombian GICOECOL. His research interests include eBusiness, eCommerce, DSL and MDA.