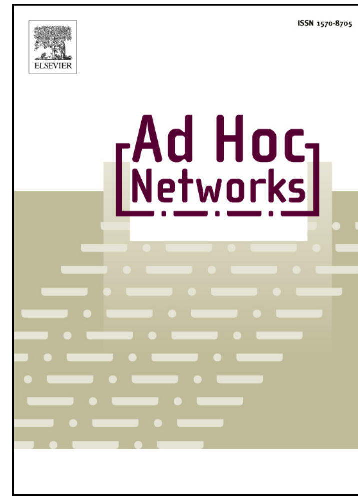
Accepted Manuscript



Onto Scalable Wireless Ad Hoc Networks: Adaptive and Location-aware Clustering

Luís Conceição, Marilia Curado

PII: S1570-8705(13)00135-2
DOI: http://dx.doi.org/10.1016/j.adhoc.2013.07.001
Reference: ADHOC 925

To appear in: *Ad Hoc Networks*

Received Date: 30 July 2012
Revised Date: 11 June 2013
Accepted Date: 1 July 2013

Onto Scalable Wireless Ad Hoc Networks: Adaptive and Location-aware Clustering

Luís Conceição, Marilia Curado

Dept. Informatics Engineering
Centre for Informatics and Systems
University of Coimbra

Abstract

Clustering is a widely used solution to provide routing scalability in wireless ad hoc networks. In the literature, clustering schemes feature different characteristics and purposes, however few schemes are context-aware. This work proposes a new solution called Distributed and Location-aware Clustering (DiLoC), a clustering scheme designed to operate in indoor environments, providing mechanisms to gather context location information in order to ease the maintenance of clusters, thus resulting in a stabler network topology in order to provide a scalable network topology for an efficient routing. DiLoC considers three distinct approaches, regarding the characteristics of the deployment environment, aiming to cover infrastructure-less, infrastructure and hybrid network scenarios. DiLoC was evaluated and compared with a similar clustering scheme, featuring the stability, amount of clustered nodes and network load. Included results demonstrate a scalable algorithm with a significant high stability.

Keywords: ad hoc networks, distributed clustering, stability, promiscuous indoor environments

1. Introduction

With the evolution of wireless technologies, there has been an increasingly wide utilization of mobile devices. Mobile networks have become particularly attractive in the recent years due to their flexibility at considerable low costs. Wireless is indeed one of the nominated communication technologies of the future, since it has the potential to allow the connection of all types of mobile devices.

Wireless ad hoc networks are autonomous systems, capable of self deployment and maintenance, not dependent from existing network infrastructures for their operation. As a result, the topology of such networks is very dynamic, especially due to the unpredictable behaviour of the nodes involved. In this context, numerous clustering schemes were developed, following different approaches and objectives, such as stability, low maintenance overhead or energy efficiency. Each one attempts to obtain the best efficiency by varying the

Email addresses: lamc@dei.uc.pt (Luís Conceição), marilia@dei.uc.pt (Marilia Curado)

characteristics of the system, like the usage of clusterheads and gateways, the maximum hop distance between nodes and the location awareness. However, there are very few clustering schemes which provide a fully distributed cluster structure with no clusterheads.

In recent years, a wide growth of wireless systems has been noticed. Wireless technologies are present in consumer applications, medical, industrial, public services, transports and much more. Therefore, there is a high demand for accurate positioning in wireless networks, either for indoor or outdoor environments. Concerning the nature of the application, different types of location are needed, which can be characterized as physical location, symbolic location, absolute location and relative location. Physical location is expressed in coordinates, identifying a point on a map. Symbolic location refers to a location in natural language, such as a coffee shop, office, etc. Absolute location uses a global shared database system, which references all located objects. Finally, relative location is usually based on the proximity of devices, e.g. known reference points, providing an environment-dependent location. The latest is the most common used paradigm in literature.

Currently, there are many wireless location technologies, such as Radio Frequency (RF) based (WLAN, Bluetooth, ZigBee, RFID), Infrared (IR), Ultrasound, and GPS. However, each technology has its advantages and disadvantages, and environment scope. No single technology is applicable to all services and circumstances. Recent studies have deeply concentrated on developing indoor location systems, since GPS offers an good solution for outdoor environments.

In this work, a new clustering scheme is proposed, namely Distributed and Location-aware Clustering (DiLoC), aiming to further improve the stability of the cluster structure.

The remaining of this document is organized as follows. Section 2 discusses the related work, covering some of most significant clustering schemes and location sensing solutions. Section 3 describes the DiLoC clustering scheme. Section 4 performs a comparison evaluation of DiLoC and, finally, Section 5 concludes the article.

2. Related Work

This section conducts a study of some important clustering schemes in literature, followed by the main location sensing systems. This overview aims to analyse the different employed characteristics and mechanisms in order to provide a better understanding of the proposed solution.

2.1. Clustering

Clustering algorithms can be classified according to different characteristics and objectives [1]. One of the common features in clustering schemes is the utilization of clusterheads (CH) and most of the proposed schemes rely on centralized nodes to manage the clusters structure. The utilization of gateway (GW) nodes is also another important characteristic that is present in the majority of clustering schemes. Other properties of clustering schemes concern the single-hop or multi-hop environments, the multi-homing (MH) support, embedded routing capabilities and location awareness. Combining the possible characteristics, each proposed clustering scheme attempts to accomplish a specific objective.

The Stable Clustering Algorithm (SCA) [2] aims at supporting large Mobile Ad Hoc Networks (MANETs) containing nodes moving at high speeds by reducing re-clustering operations and stabilizing the network. To meet these requirements, the algorithm is based on the quick adaptation to the changes of the network topology and reduction of clusterhead reelections. In order to avoid a high frequency of clusterheads reelection, the algorithm initially chooses the nodes that best meet some required metrics such as, energy, mobility, connectivity and communication range. The Enhanced Sectorized Clustering Scheme based on Transmission Range (ESCS) [3] also pays attention to the connectivity and energy of nodes. Clusterhead nodes are chosen according to the energy amount of nodes, selection those with potentially longer lifetime. For this election, the connectivity of nodes is also accounted by measuring the node density, i.e. the amount of nodes within a sector. Results feature the network lifetime, showing that this schema is superior, particularly for scenarios with a larger amount of nodes. The Signal Energy Efficient Clustering (SEEC) [4] is also based on the energy level of nodes and signal strength. Once more, the these metric are used particularly in the election of clusterheads. SEEC constantly monitor the energy levels of clusterheads and preventively replace them before energy is depleted.

The Stability-based Multi-hop Clustering Protocol (SMCP) [5] also builds the cluster structure according to the node connectivity quality. Moreover, this scheme introduces a new methodology (clustercast mechanism) with the purpose of limiting the broadcast of less significant control messages. The K-hop Clustering Protocol (KhCP) [6] protocol is specifically designed to cluster dense MANETs, as it delimits the cluster formation at a specified k-hop distance. In this protocol, clusters are formed on a circle basis, whereas the clusterhead, at the start point, is the centre of the circle. A weight-based clustering scheme, named Distributed Weighted Clustering Algorithm (DWCA), was proposed with the objective to extend the lifetime of the network, by creating a distributed clustering structure [7]. The election of clusterheads is based on the weight value of nodes, which is calculated according to their number of neighbours, speed and energy. The Enhanced Performance Clustering Algorithm (EPCA) [8] is also a weight based clustering solution. Once more, the weight parameters are only taken into account for the selection of the clusterhead.

The Connectivity-based Clustering Scheme (CCS) [9] has the purpose of improving the effectiveness, reliability and stability of MANETs. In contrast with most schemes, this solution ignores mobility and energy parameters, focusing only in the cluster organization to achieve its objectives. In order to provide effectiveness and low maintenance, it utilizes a technique of maintaining clusterheads separated by a significant hop distance. Therefore, the probability that two clusterheads come into each other's transmission range is reduced, decreasing the number of re-clustering operations. Concerning the reliability objective, an intra-connection degree is used to measure the connection quality between a node and the possible clusters that it can join. The Energy Efficient Mobility-sensitive Clustering (EEMC) [10] presents a solution for energy balancing. The main objective of this scheme is to extend the lifetime of the network, by distributing the load amongst nodes and also regarding their mobility. The Trust-related and Energy-concerned Distributed MANET Clustering (TEDMC) [11] is also a scheme driven by energy concerns. TEDMC considers that the most important nodes are the clusterheads, and therefore it elects them according to their trust level and residual energy. In order to keep information about the trust level of nodes, this algorithm maintains and periodically exchanges a reputation rank table, which contains a reputation value and the unique identification of the last node to assign the value in question. Furthermore, TEDMC is substantially different from KhCP, as it only allows 1-hop clusters, thus being less suitable for dense networks.

Table 1: Comparison of clustering schemes

	CH	**GW**	**Hops**	**MH**	**Main Objective**
SCA [2]	Yes	Yes	2-hops	No	Large MANETs with high-speed nodes
ESCS [3]	Yes	Yes	n-hop	No	Optimal clusterhead election for improved stability
SEEC [4]	Yes	Yes	n-hop	No	Longer network lifetime based on clusterhead energy
SMCP [5]	Yes	Yes	n-hop	No	Stable cluster formation
KhCP [6]	Yes	No	n-hop	Yes	Limited overhead for dense networks
DWCA [7]	Yes	Yes	1-hop	No	Stability of the network
EPCA [8]	Yes	No	n-hop	No	Performance, with trusting node mechanism
CCS [9]	Yes	No	n-hop	No	Effectiveness, reliability and stability
EEMC [10]	Yes	No	n-hop	No	Distributed power consumption, limited control message flooding
TEDMC [11]	Yes	Yes	1-hop	No	Stability, relying on trust values and residual energy of nodes
OCRP [12]	Yes	Yes	1-hop	No	Merge clustering phase with routing discovery and data transmission
ORC [13]	Yes	Yes	n-hop	Yes	Light control overhead, providing cluster structure and routing
ODGM_GN [14]	Yes	No	n-hop	No	Build clusters as foundation for variable types of routing protocols
EWDCA [15]	Yes	No	n-hop	No	Maintain stable cluster structure with lowest number of clusters
NSLOC [16]	No	Yes	n-hop	No	Provide stable cluster structure with low overhead
AMC [36]	Yes	Yes	n-hop	No	Topology changes with low overhead for highly mobile networks
SALSA [37]	No	Yes	n.hop	No	Low maintenance overhead with balanced clusters

There are also clustering schemes capable of performing route discovery, such as the On-Demand Clustering Routing Protocol (OCRP) and On-Demand Routing-based Clustering (ORC) [12, 13]. These schemes are capable of building cluster structures and routing paths on-demand. In these schemes, only the nodes that are necessary to satisfy a routing path are bounded to the cluster structure. The On-Demand Group Mobility-Based Clustering with Guest Node [14] provides a solution with the main purpose of building a cluster structure capable of supporting several types of routing protocols with identical efficiency. Furthermore, it relies in a guest node approach in order to introduce arriving nodes to the network.

The Efficient Weighted Distributed Clustering Algorithm (EWDCA) [15] has the major concern of providing scalability for MANETs, by taking into consideration several weight parameters: connectivity, residual battery power, average mobility and distance between nodes. These parameters are used only to elect the most suitable clusterhead, in order to keep an optimal number of clusters, thus providing as much scalability as possible. A Novel Stable and Low-maintenance Clustering Scheme (NSLOC) [16] is a fully distributed clustering scheme, with the main goal of simultaneously providing a low maintenance overhead and network stability. The NSLOC scheme employs a completely distributed approach, not relying on clusterheads, in contrast to most well known clustering schemes. The Adaptive Multihop Clustering (AMC) [36] scheme is designed to operate in highly mobile networks, providing a low cluster management overhead. To achieve this objective, the AMC is relies on an adaptive management mechanism, keeping clusters with an even topology. It is based in a lower bound and an upper bound, forcing the clusters to have minimum and maximum size, respectively. When the size of a cluster is less than the lower bound it is merged with a neighbour cluster, but only if their combined size is less than the upper bound. On the other hand, if a cluster size overcomes the upper bound, it is divided in two separate clusters. The Smart and Balanced Clustering for MANETs (SALSA) [37] is a distributed scheme (clusterhead free), also providing a balanced clustering mechanism to maintain an even topology with a low maintenance overhead. The even topology is achieved with a load balancing mechanism. When the size of a cluster is becoming close to the maximum allowed, it begins to reassign nodes to a neighbour cluster. Likewise, when a cluster size is becoming low, it also begins to reassign nodes from a neighbour cluster. This balancing mechanism is more progressive than the one used in AMC, since its does not radically change the entire cluster structure. SALSA also transmits only the necessary information meeting specific requirements, avoiding unnecessary overhead.

Table 1 shows the main characteristics of the analysed clustering schemes. One of the main reasons clusterheads are so utilized is due to the simplicity that they provide to the clustering algorithm. Centralizing the management power on only one node results in a less complex algorithm thus, becoming easier and faster to implement. Nonetheless, clusterheads carry big disadvantages, as they represent bottlenecks and uneven energy consumption in the network, due to the centralized management decisions.

2.2. Location Sensing Systems

The main challenge of location estimation relies on the radio propagation interferences, due to severe multipath, low probability of a Line-of-sight (LOS) path, reflecting surfaces, and environment dynamic characteristics, such as building restructuring and moving objects. There are three main techniques to model radio propagation: trilateration, fingerprinting and proximity.

- *Trilateration* - is the most used technique by which the location of devices can be determined. The process consists on determining radial distance, obtained by the received signal, from three or more different points. Trilateration can be used on most RF based technologies by measuring radio propagation characteristics, thus calculating distances from two different points. If the position of three access points A, B and C and the distances of MA, MB and MC are known, it is possible to obtain the M relative position by the trilateration method.

- *Fingerprinting* - In contrast with Trilateration, RF based fingerprinting algorithms first collect features (fingerprints) of a scene and then estimate the location of devices, by matching (or partially matching) real time (online) measurements with fingerprints. Most of these algorithms define location fingerprints based on RSS values, previously obtained (offline). Thus, the fingerprinting technique must occur in two stages: the offline gathering of fingerprints, where multiple measurements of known locations are stored in a database, and a online location estimation, which obtains the most suitable match from the database. The major challenge of this technique is the dynamic environments, since building layouts and arrangement of objects are likely to change, thus affecting RSS measurements.

- *Proximity* - algorithms determine symbolic locations. Typically, it relies on the installed base stations, each classified to be in a known position. When a mobile device is detected by the BS antenna, it is considered to be located in its coverage radius. Moreover, when multiple antennas detect a device (overlapping), it is considered to be located in the BS with the strongest signal, for which the RSS value is typically used. This technique is simple to implement and it offers reasonable results in Radio Frequency Identification (RFID) and IR based algorithms, due to their low range.

There are many proposed wireless location solutions, using different technologies, scopes and with different accuracies. Two main approaches are used, when developing a location system. The first contemplates the usage of existing infrastructures, such as access points, and used them to locate devices. The second approach consists in the development of the entire system, the network infrastructure and the signalling system. In the first, installation costs are low and usually, the system can be deployed on different scenarios, without additional costs. On the other hand, the development of a new system, only designed for location is favourable, potentially increasing the positioning accuracy. In this document, pioneer and recent solutions are discussed, regarding their technology, accuracy, range and dimension capabilities. Table 2 shows the analysed location sensing solutions.

Infrared (IR). IR based technology provides advantages such as the restriction of signals within rooms (IR) beams do not pass through walls) and immunity to Electromagnetic Interference (EMI), in contrast with RF based technologies.

Table 2: Wireless-based Location Sensing Solutions

Solution	Technology	Accuracy	Indoor range	Dimension
Active Badge [17]	IR	Exact room	< 6m	2D
LocSens	RF	1.05m - 2.90m	< 20m	2D
COMPASS	RFID	m	m	2D
RADAR [18]	WLAN	2m - 3m	25m - 50m	2D
HLPS [19]	IR + RF	Exact room (indoor)	< 8m	2D
LANDMARC [20]	RFID	1m	50m	2D
ZRL [21]	ZigBee	Exact room	10m	2D
Bat [22]	Ultrasound	0.03m	6m	2D
Cricket [23]	Ultrasound + RF	0.06m - 0.15m	6m	2D
3DIL [24]	Bluetooth	2m	6m - 10m	2D/3D
ILB [25]	Bluetooth	2m - 4m	6m	2D
AMRL [26]	Bluetooth	0.198m - 0.656m	10m	2D
WLD [27]	WLAN	3m - 4m	50m	2D
ILT [28]	WLAN	1m - 4m	25m	3D
A-GPS [29]	GPS	5m - 50m	NA	3D
RTLS [30]	GPS + RFID	-	-	2D/3D

Furthermore, the signal power can be adjusted in order to cover only small areas.

The Active Badge [17] system was a pioneer contribution in location sensing systems and source of inspiration to many following projects. The main goal of this solution is the ability to locate persons or objects inside public buildings like hospitals. Each person wears a badge, which emits an IR signal within every 10 seconds. The sensors placed at known positions are responsible to receive the unique identifiers and relay these to the location manager software. Emitted signals are reflected by surrounding materials and therefore are not directional when used inside small rooms. Pfeifer *et al.* [19] proposed an hybrid IR/RF solution suitable for outdoor and indoor location tracking. Each user wears a IR receiving and RF transmitting badge. The scenario is equipped with stationary, stand-alone infrared ID beacons and with stationary RF receivers with LAN (or WLAN) interface. In contrast with Active Badge system, IR transmitters are scattered through the scenario to reduce the cost of badges, as receivers are around 10 times cheaper. Upon receiving a new IR ID beacon, badges transmit a message to the base station via RF, informing the management system of its location. At outdoors, the sunlight interferes with the IR transmission. In this case, the badge is localized relatively, by proximity measurements using the Receiver Signal Strength Indication (RSSI), to the base station.

WLAN (802.11). The WLAN standard has become widely spread across public hotspots, corporative locations and many types of mobile devices. The dominant role of this technology appeals to the utilization of existing infrastructures for indoor and outdoor location sensing. The typical accuracy of WLAN positioning systems using RSS is approximately 3 to 30 meters.

Bahl *et al.* [18] proposed an WLAN indoor location tracking system called RADAR. In this work, two main types of approaches are employed to determine user location: empirical model and radio propagation model. The first depends on a database that consists of previously measured signal strength of points, recording user orientation and signal strength for each Base Station (BS). In the second approach, authors adopted the Floor Attenuation Factor (FAF) and Wall Attenuation Factor (WAF) models [31], taking into consideration the number of obstructions walls and material types between the user and the BS. These values depend on the building layout and must be derived empirically. The accuracy of RADAR is approximately 2 to 3 meters. Despite the low installation cost, RADAR has a big disadvantage. The collection of reference data is always necessary for both approaches. Each change in a room structure requires an update of the reference model. Sanchez *et al.* [27] presented a location determination method which limits search area to a starting point, inferred by typical trilateration. Upon the trilateration of a device to a point, search will be constrained to a surrounding area (search are reduction), thus decreasing computation cost. Results show an acceptable accuracy where 90% of the cases is within three meters. Hossain *et al.* [28] proposed a 3D location tracking system, using the trilateration technique based on RSSI values, as well. The testbed scenario consisted on an area with 12 meters x 12 meters of size with two floors, whereas three Access Points (APs) were attached at fixed positions. During tests, mobile nodes moved freely, measuring RSSI values and sending them to the central processing server, which is responsible to calculate locations.

Bluetooth (802.15). Compared to WLAN, the Bluetooth gross bit rate is lower (around 2 Mbps max.), and the range shorter (typically 10 - 15 meters). On the other hand, Bluetooth technology is lighter standard, providing a less complex stack and support to other networking services in addition to IP.

Rodriguez *et al.* [25] proposed a new indoor location system based on the Bluetooth technology. Access points of the network are used to provide network access and location estimates. Nodes measure the RSS values received from access points and sent them to a central server, through the network, where location calculations are performed. Since processing power is concentrated in a powerful/wired up machine, it is possible to use any kind of location estimation algorithm. Furthermore, the location system uses a reference scene analysis of signal strength, previously performed. This approach is similar to RADAR [18], where RSSI readings are sent from mobile nodes to the server in tuples.

Raghavan *et al.* [26] proposed an location system, for indoor environments, suitable to any technology that provides RSSI values, such as Bluetooth and WLAN. However, since it is designed to locate robots, the authors chose to use Bluetooth, as power consumption is significantly lower than WLAN, despite of providing a higher data rate. This approach uses a different trilateration method, namely iterative trilateration, as employed in [32]. The method can provide more accurate results, however at a higher processing cost, by discarding the points with a low error, and repeating the computation process to the remaining.

More recently, Cruz *et al.* proposed a 3D indoor location and navigation system using Bluetooth radio technology [24], implemented using Java and J2ME. Location calculation is performed using the kNN (k-Nearest Neighbors) [33] method and its conducted by the mobile device.

RFID. RFID is a technology capable of storing and transmitting data to an RF compatible circuit. Typically, these systems are composed by RFID readers and RFID tags. The RFID reader is able to read the data emitted from RFID tags, using a defined RF and protocol exchange information. Moreover, RFID tags can be either passive or active. Passive RFID tags operate without a battery and require stronger signals from the reader. They are much lighter, smaller and less expensive than active tags. Basically, passive tags reflect the RF signal transmitted to them from the reader and add information by modulation the reflected signal. The signal strength returned from these tags is, however, constrained to very low levels (readers power the tags), ranging from 1 up to 3 meters. Active RFID tags are small transceivers, which actively broadcast data, such as their ID, in response to an interrogation. The advantages are clearly the transmission range and the low required signal strength emitted by readers.

LANDMARC [20] is an indoor location sensing system using active RFID, aiming to locate objects. The infrastructure consists of RFID readers, active RFID tags and a management server. All objects must be tagged with an active tag. Active tags are also deployed across the scenario, acting as reference tags, aiding the location process with a low installation cost. This approach requires signal strength information from each tag to readers, whereas location estimation is performed using the kNN (k-Nearest Neighbors) [33] method. The server communicates with the readers to receive RSS measurements and calculates the estimated positions of targets. The main disadvantage of this approach resides in the sequential scan of all reading ranges, which takes about one minute per cycle.

Cheng *et al.* proposed a system to improve localization accuracy of objects in hospitals for health and safety monitoring, namely COMPASS [34]. The innovation of the COMPASS algorithm is the ability to estimate the position of tracking tags based on a cluster (community) of four reference tags. Rather than trying to compute the position of the tracking tag based on each individual reference tag, a community based localization method is employed to improve tracking accuracy. This approach has more accuracy and lower error when compared to LANDMARC results. However, a larger quantity of RFID readers, acting as reference tags, are needed.

Zigbee (802.15.4). ZigBee is an RF wireless technology designed for small devices, including several benefits such as low power consumption, simplicity of the stack, and easy deployment on Wireless Sensor Network (WSN).

Cheng [21] proposed a room-based location technology using ZigBee wireless technology. Two ZigBee nodes are placed inside each room, one at the door, with the antenna pointing inwards the room and adjusted within 1.5 meters, and a second in a unspecific wall, adjusted within 10 meters. When the user tag passes the door or room and the secondary node senses the user tag, it can be certain that the user is in that specific room. The node positioned at the door is crucial as it avoids location miscalculations when the range of secondary nodes overlap. This approach is quite different from the majority of proposals in literature, since it does not need to calculate location based on RSS information, therefore requiring a very low processing power.

Bras *et al.* [35] proposed a ZigBee location protocol based on WSN. The main objective of this approach is to reduce power consumption. To achieve that, authors constrained the traffic exchanged by the coordinator and used a custom built routing protocol. The coordinator in the core processing module of the network, which receives the RSSI values from mobile nodes in the ZigBee network and calculates its relative location. Furthermore, the routing protocol provides two modes of location: HiRSSI mode is a macro-based location, without need for calibration, ready to use from the start up. However, if more precise location is needed, multi-RSSI mode supports more complex algorithms based on RSSI analysis, collecting RSSI values from several devices.

GPS. The Global Positioning System (GPS) is one of the most successful position systems for outdoor environments. Thus, the discussion of this system alone, is somehow futile. There are, however, hybrid systems than combine GPS with other technologies, capable of providing indoor and outdoor location sensing.

Misra *et al.* pioneered in an hybrid GPS solution, called Assisted-GPS (A-GPS), providing outdoor and indoor positioning with an accuracy of 5-50 meters. A-GPS technology uses a server with a reference GPS receiver, improving a partial GPS receiver find weak GPS signals. The references from the database can be obtained via internet connection, collecting the necessary information according to the wireless mobile network.

Guillemette *et al.* [30] proposed an hybrid RFID-GPS location system, applied to a real scenario, whereas indoor and outdoor positioning is performed with the RFID and GPS technologies, respectively. The system was designed to monitor security guards patrolling a campus. For outdoor tracking, GPS receivers were integrated with radios, which periodically report its position using an defined RF. In indoors, RFID tags were installed in buildings, able to receive beacons sent by the RFID beacons, also installed on the radios. When RFID beacons receive responses to the emitted signals, a message is sent via WLAN technology to the WLAN base stations, also installed in the campus. With this system, it is possible to determine, in real time, the location of guards. Authors do not discuss energy concerns, however, since each radio is equipped with a WLAN card, a GPS receiver and an active RFID, the devices are likely to have low autonomy.

Ultrasound. The most important property of ultrasonic location systems is that they have the capability to be fine-grained, meaning that it is possible to estimate location with a high degree of accuracy. This occurs because the speed of ultrasonic waves in the air is sufficiently slow to allow the TOF of a signal to me accurately measured between sender and receiver.

The Bat system [22] presents a location approach based on ultrasound. Each person or object carries a device called Bat that periodically sends an ultrasonic signal. Receivers are placed to fixed positions at the ceiling of rooms, and connected to a wireless network. Analyzing the arriving times, provided by several receiving units, the core management system calculates the position of devices. This project shows that ultrasound provides an high precision location sensing, however ultrasound is highly vulnerable to interferences.

Cricket [23] is another ultrasound based location system. In contrast with the Bat system, mobile devices are responsible to determine the location by themselves, ensuring privacy to the users. Also, instead of receivers, beacons are placed in the ceiling, and periodically send radio and ultrasonic signals. Using multiple signals from different beacons, the mobile device calculates the current position.

2.3. Summary

This subsection discussed some of the most important positioning techniques and solutions. Different technologies were presented, exhibiting the main trade-offs between them. It was observed that the main challenge of location sensing lies in indoor environments, since GPS technology is capable of providing good results in outdoors. Currently, there is not a location solution, using a single technology, capable of providing indoor and outdoor coverage. Different technologies imply a different granularity and accuracy of the location system. Thus, several proposals are hybrid, compiling multiple technologies. However, these solutions have energy issues, due to the necessity of multiple hardware parts to support all the employed technologies.

3. Distributed and Location-aware Clustering (DiLoC)

Location awareness is a key feature in distributed networks, particularly for clustering schemes. With location sensing mechanisms, each node can join and change clusters more efficiently, with less overhead. Since nodes are assigned to clusters more rapidly, the amount of time within clusters will increase, therefore improving the stability of the network.

The great challenge of location sensing systems features mainly indoor environments, since GPS is a fair acceptable solution for outdoors.

This work proposes DiLoC, a new clustering scheme based on indoor location, designed to operate in infra-structureless and infrastructure network environments. As studied in section 2.2, location sensing systems are complex, particularly when concerning trilateration and fingerprinting. Typically, in wireless ad hoc networks it is not feasible that each node computes its location using triangulation, otherwise it would result in a very slow system, with overhead issues. Furthermore, fingerprinting is also not suitable, since nodes would have to query the database for their location (resulting in a large overhead), and this type of networks are usually deployed in unknown scenarios.

DiLoC is based on proximity location, relying on devices scattered along the network to determine location information for the entire cluster structure. Generally, there are two distinct types of methodology in order to obtain proximity location. In the first methodology it is assumed that the deployment scenario does not have network infrastructures (e.g. WLAN network), therefore it is necessary to pre-elect special nodes to provide location awareness. The second methodology takes advantage of the existing WLAN network infrastructure, utilizing its Access Points (APs) as references to ease the construction and maintenance of clusters.

The DiLoC scheme contains three different approaches regarding the target deployment scenario: No Infrastructures (NI), Basic Infrastructures (BI) and Advanced Infrastructures (AI). In the first approach the algorithm does not rely on any network infrastructures to retrieve location information. The second approach only relies on the existing network infrastructures. Finally, the third approach is designed to operate in hybrid scenarios, where network infrastructures are not certain to be present.

3.1. General Algorithm Description

The main purpose of DiLoC is to build a low overhead clustering structure, aiming to increase network stability with node location awareness. It is designed to build a cluster topology in a distributed fashion (contrary to the clusterhead paradigm), providing a light hierarchical structure for routing. DiLoC is a new clustering scheme, implemented from scratch, which extends some of SALSA [37] features, such as the load-balancing mechanism and the structure of packets (detailed further in this document). The load-balancing mechanism, which acts progressively during execution, distributes nodes across clusters based on their current size. Before the maximum capacity of a cluster is reached, nodes are assigned to a neighbour cluster or, in cases this operation is not possible, builds new clusters to receive excess nodes, keeping an even topology. The DiLoC algorithm is based on state transitions, to ease the difference of operations.

3.1.1. Node States

Nodes can be in one of three distinct states, namely *Unclustered*, *Clustered* and *Clustered-GW*, as shown in Figure 1. The current state of a node strongly depends on being in-range to its neighbour nodes. Two nodes are considered to be in-range to another when they are able to exchange information. In DiLoC, a *Ping* message (further explained in Subsection 3.1.3) is used as a beacon to determine if two nodes are in-range.

The *Unclustered* state typically represents a temporary role, as the node is waiting to be assigned to a cluster. In this state, when the node discovers neighbours, it waits a predefined period of time in order to calculate the best candidate cluster to join.

Nodes in the *Clustered* state usually represent the majority of nodes on the network, whereas all in-range nodes must belong to its cluster. Thus, the communication with foreign nodes (i.e. nodes assigned to a different cluster) is performed through gateway nodes.

Finally, the *Clustered-GW* state is assigned to nodes that have in-range foreign nodes, i.e. they must have direct connectivity with at least one different cluster. Thus, they are responsible of forwarding inter-cluster maintenance messages and typically are located on the edge of clusters.

3.1.2. State Transitions

The *Unclustered* state occurs on two different situations:

1. Node isolation - in this case the node does not have any in-range neighbour nodes, therefore cannot create or be assigned to a cluster
2. Cluster transition - the management of clusters occasionally requires nodes to change clusters, due to cluster balancing. In this phase, nodes can be unassigned from a cluster.

Unclustered *to* Clustered. This state occurs when a node becomes aware of an in-range cluster or an unclustered node. In the first situation, the node joins the cluster automatically. However, if the node only detects unclustered nodes, a new cluster is created to adopt the unclustered nodes.

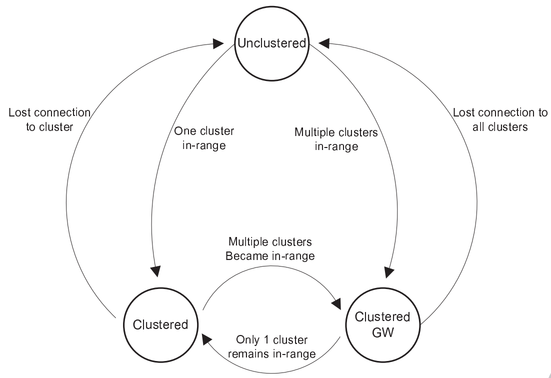


Figure 1: Node states

Unclustered *to* Clustered-GW. This transition is similar to the previous, but more than one cluster is discovered. Firstly, the node calculates which is the best, taking several parameters into account: number of in-range nodes for each cluster and the size of clusters. The greater the number of in-range nodes, the stronger connection to the cluster. However, if the size of the cluster is high, possibly close to the maximum allowed, this cluster would be a bad choice. To measure this trade-off, a new metric is utilized (1), namely the best clustering metric (BC), where BC_i is the metric value for cluster i.

$$BC_i = AP_i + \frac{IRN_i}{C} \tag{1}$$

AP_i is defined as the number of the available positions in cluster i until it reaches the maximum allowed, i.e. the difference between the maximum allowed number of nodes per cluster and the current number of assigned nodes. IRN_i is the number of in-range nodes belonging to the cluster. C represents a constant value, allowing IRN_i to be less relevant than AP_i, since the number of available positions in a cluster is typically more important than the number of in-range nodes. Thus, the C value should be chosen according to the target scenario. The clusters in scenarios with a high node density have the tendency to be full, thus the C value must be large to provide more relevance to the available cluster positions. On the other hand, in small density networks, clusters have the tendency to be less populated, and a low C value should be chosen, providing more relevance to nodes in transmission range. The cluster with the higher BC value is chosen by the node.

Clustered *to* Clustered-GW. This transition occurs when a node becomes aware of clusters, excluding its own.

Clustered-GW *to* Clustered. Whenever a clustered gateway node loses connection with all its foreign clusters, it automatically transits to a normal clustered state.

Clustered/Clustered-GW *to* Unclustered. A node becomes unclustered when willingly disconnects from the network or loses connection with all its neighbour nodes. When this situation occurs, it is necessary to verify the consistency of the cluster, i.e. guarantee that all home nodes can communicate with each other.

Table 3: Node maintenance information

Information	Description
Node_ID	Unique identifier of the node
State	Current state of the node (*Unclustered*, *Clustered* or *Clustered-GW*)
C-Degree	Value to determine the connection type towards this node. Value ranges from 0 to 5, whereas 0 represents a non-neighbour (therefore merely a home node), 1 denotes a lost connection towards this node and finally, 2-5 values represent the quality of the connection, being 5 the best possible connection.
Alive	Boolean value, determining whether the node is responding or not

3.1.3. Maintenance Information and Messages

This subsection describes the information that each node maintains and the messages utilized. There is a table providing insight of the network topology, namely the *NODE_TABLE*. This table keeps all the information about neighbour and home nodes, as described in Table 3.

DiLoC relies on multiple, small purpose-driven messages to manage the cluster structure. All messages contain one common field, *Type_ID*, which uniquely identifies the message type that is being transmitted. Apart from this field, all the messages contain different sets of fields, suitable to their purpose, as follows:

- *Ping* - periodic broadcast message, allowing nodes to discover their neighbourhood.
- *Hello* - provide the structure of the cluster to member nodes.
- *Lost Hello* - broadcasted when a node loses connection with a neighbour home node, informing member nodes, that do not have direct connection, about a possible disconnected node. This event triggers a process in order to verify if the node is still connected via other nodes, namely the alive check process. At the end of this process, if it is verified that the node is in fact disconnected, it is necessary to verify if the cluster is still consistent, which implies the utilization of the following described message (*Alive Hello*).
- *Alive Hello* - upon the trigger of an alive check process, to verify the consistency of the cluster, i.e. guarantee that all nodes inside the cluster are capable of communicating with each other. In most situations the cluster remains consistent; however there are cases in which the cluster becomes partitioned in two clusters. In this particular situation, both clusters have the same identifier, thus it becomes imperative to change it.
- *Switch Hello* - used when a cluster identifier becomes inconsistent and it is necessary to change the *Cluster_ID* for their nodes.

3.2. Description of DiLoC Aproaches

As briefly described, the new algorithm implements three distinct approaches to handle location proximity. The first assumes that the deployment scenario

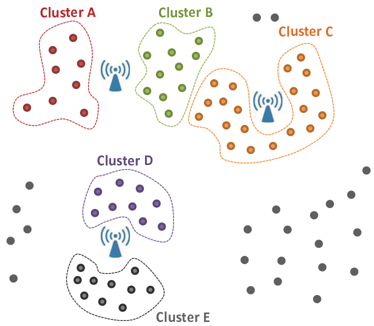


Figure 2: Hypothetical Scenario in BI Approach

does not have a WLAN network infrastructure and therefore, it uses special nodes serving as reference location points. The second approach is based on an existing WLAN network infrastructure, taking advantage of its APs as reference, in order to provide location sensing. The third is a combination of the previous two approaches, aiming to provide an hybrid solution, suitable to the majority of scenarios.

3.2.1. First Approach - No Infrastructures (NI)

This algorithm is designed to operate in indoor environments, where network infrastructures are not available. Therefore, in order to manage clusters based on location information it is necessary to employ the following described technique. During the initial period of network deployment (e.g. network cold start), some configured nodes are assigned to a special role, namely anchor nodes, which will serve as location reference points. Anchor nodes are deployed in a scattered fashion in order to cover as much as area as possible, and remain static for a certain period of time. These nodes are responsible to immediately create clusters and announce them using appropriate messages. This process, accelerates the initial topology setup, in contrast with the traditional exchange of messages employed in solutions designed for MANETs [37]. Once an anchor node detects that regular nodes became assigned to its cluster, it becomes a regular node itself, being able to move and disconnect, i.e. following the exact same rules of regular nodes. Anchor nodes are a complementary tool in order to create clusters more efficiently and with an initial balanced distribution of nodes. Moreover, anchor nodes may be non-existent or disconnected at any moment. If such is the case, regular nodes are also able to create clusters, however in random locations, causing the initial cluster topology to be less evenly distributed.

3.2.2. Second Approach - Basic Infrastructures (BI)

In contrast with the previous approach, this algorithm is designed to operate in scenarios where location information can be retrieved using the WLAN network infrastructure.

If an wireless ad-hoc network is to be deployed in a scenario where a WLAN network is present with APs' coverage (e.g. university campus), the overhead associated with the distribution of anchor nodes can be avoided. This approach is capable of recognizing APs as relative positions, around which clusters can be created and maintained. The obvious drawback is the strong dependence of an existing infrastructure. In the initial network deployment, nodes scan for WLAN SSID broadcasts and create clusters according to that information, requiring less message overhead than the previous approach. Nodes around APs are instantly assigned to a cluster according to the SSID string, whereas the initial waiting period for cluster creation is not necessary. Furthermore, all nodes have the same behaviour, not being necessary to statically configure anchor nodes prior to network deployment. When an unclustered node receives an SSID broadcast, it must first analyse if a neighbour cluster exists, i.e. if there are in-range nodes already assigned to a cluster. If a cluster is present, it is immediately assigned to that cluster. Otherwise, it analyses if there are any in-range unclustered nodes. If other nodes are present, it then creates a new cluster with a globally unique identifier and broadcasts that information, announcing the existence of a new cluster. However, if the node is älone;, it does nothing, remaining unclustered. In the best case scenario, each broadcast would be associated with one single cluster, however this situation is not always possible. Looking at Figure 2, Cluster A and B are associated with one specific SSID AP broadcast and they cannot be merged into one single cluster, since its nodes do not have connectivity. The same situation occurs with Cluster D and E.

3.2.3. Third Approach - Advanced Infrastructures (AI)

Previous subsections describe two different approaches of the DiLoC algorithm, which contemplates scenarios with and without a network infrastructure. However, a combined approach must be considered, since there are hybrid deployment scenarios, i.e. one part of the scenario has a network infrastructure and the other part does not. Thus, the AI based approach aims to provide a cluster topology covering all areas of the scenario. This approach is a combination of the previous two, with each node operating accordingly to its situation. It is capable of both exploiting network infrastructures as well as creating and managing clusters for isolated nodes (i.e. when nodes are not in-range from the remaining cluster structure). Figure 3 shows a hypothetical scenario where APs are deployed. As depicted, clusters are built around APs and isolated nodes are also clustered, in contrast with the BI approach. Since it combines two different paradigms, nodes have to decide which approach to use for different situations. There is a basic rule that defines the behaviour of a node: if it receives broad-

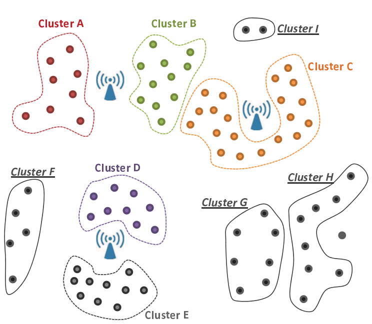


Figure 3: Hypothetical Scenario in AI Approach

casts from an AP or from an existent cluster, it will belong to that cluster. Otherwise (i.e. it is isolated) it will wait for the presence of surrounding clusters, potentially created by anchor nodes. To be noted that should an anchor node be in-range to an AP, the node will immediately resign from that role and become associated with the cluster near the AP.

3.3. Algorithm Specification

This subsection illustrates the main structure of the algorithm execution for the unclustered and clustered states of nodes. The three approaches of DiLoC mainly differ in the unclustered state, whereas each follows different rules for cluster creation.

3.3.1. Unclustered *state*

In the unclustered state only two types of messages are accepted, namely *Ping* and *Hello*. *Ping* messages are used to update the status of nodes and are the only to trigger cluster creation and joining events. *Hello* messages are only used to update the cluster member nodes table. DiLoC acts differently according to the configured approach. The main differences in the algorithms are defined in the following pseudocode.

Algorithm 1 Received Message - NI Approach

```
if rx_Ping then
    update_tables();
    if sender node is clustered then
        start cluster_join_process();
    else if this is anchor node then
        start cluster_create_process();
        broadcast_ping();
    end if
else if rx_Hello and sender node is clustered then
    update_tables();
end if
```

Algorithm 2 Received Message - BI Approach

```
if rx_Ping then
    update_tables();
    if sender node is clustered and sender_SSID == this_SSID then
        start cluster_join_process();
    else
        start cluster_create_process();
    end if
else if rx_Hello and sender node is clustered then
    update_tables();
end if
```

Algorithm 3 Received Message - AI Approach

```
if rx_Ping then
    update_tables();
    if sender node is clustered then
        start cluster_join_process();
    else if (anchor_node and this_SSID==NULL) or (this_SSID != NULL) then
        start cluster_create_process();
    end if
else if rx_Hello and sender node is clustered then
    update_tables();
end if
```

3.3.2. Clustered *state*

In the clustered state, all types of messages are accepted. Once a clustered node receives a *Ping* message from a clustered node, it must check if its

NODE_TABLE contains the sender node. If it does not, it means that the sender node is a new member of the cluster structure, and a *Hello* message must be sent, in order to provide the information about its home nodes.

Algorithm 4 Received Message - NI, BI and AI Approaches

```
if rx_Ping then
    update_tables();
    if sender is new clustered then
        broadcast_hello();
    end if
else if rx_Hello then
    if own_cluster_id == sender_cluster_id then
        update_tables();
        forward_hello();
    end if
else if rx_Lost_Hello then
    if own_cluster_id == sender_cluster_id then
        forward_lost_hello();
        node_table_set_all_dead();
        broadcast_alive_hello();
        start_alive_timer();
    end if
else if rx_Alive_Hello then
    forward_alive_hello();
    if own_cluster_id == sender_cluster_id then
        node_table_set_all_dead();
        start_alive_timer();
    end if
    set_sender_alive();
else if rx_Switch_Hello then
    if own_cluster_id == sender_cluster_id then
        update_tables();
        own_cluster_id = sender_new_cluster_id;
        forward_switch_hello();
    end if
end if
```

The remaining received messages are only accepted if sent by a home node, i.e. a node belonging to the same cluster. If an *Hello* message is received, the maintenance tables are updated and the message is forwarded. If a *Lost* message is received, it means that some node in the cluster lost connection with another. In this case, each receiving node considers all nodes in the cluster as dead, and only are considered alive upon the reception of an *Alive* message from each one. Thus, when a *Lost* message is received, all nodes in *NODE_TABLE* are set to dead and a request to check their status is sent, by the broadcast of an *Alive* message. Likewise, upon the reception of an *Alive* message, the sender node is set to alive. This whole process is required for failsafe purposes, since in some cases the lost of a home node is vital for the cluster stability. For instance, if a cluster has the geometric shape of an hourglass, in which only one node is located in the middle, connecting the cluster, the cluster becomes incoherent if it looses connection with that particular node. When a node receives a *Switch* message, the node rapidly changes its cluster to the one specified in the message and the maintenance tables are updated according to the information received. Also, in this case, the message is forwarded for remaining nodes.

4. Simulation Evaluation

To properly examine DiLoC, a simulation evaluation, driven by its main objectives was performed, using the OPNET Modeler [38]. Thus, the main purpose of this simulation evaluation is to assess the stability and low overhead capabilities of the proposal. To accomplish this objective, a set of different simulation environments, featuring the network size and speed of nodes, were defined.

This evaluation compares the three different approaches of the proposal (No Infrastructures [NI], Basic Infrastructures [BI] and Advanced Infrastructures [AI]) alongside with SALSA algorithm, in order to capture the advantages and drawbacks of the algorithms.

4.1. Environment and Parameters

The performance of clustering schemes is strongly influenced by the scenarios under which they are evaluated. For instance, a better performance is expected for low-density networks (i.e. low quantity of nodes per Km^2) or with nodes moving at low speeds. The evaluation settings were selected in such a way that they represent, as much as possible, realistic scenarios. In this specification the evaluation parameters can be divided in three groups, the fixed-value, the variable-value parameters and the DiLoC specific parameters (Table 4).

Table 4: Simulation parameters

Fixed-value parameters	
Simulator	OPNET Modeler 16.0
Field Size (m^2)	500×500
Node mobility algorithm	Random Waypoint Model
Pause time (s)	50
Transmission range (m)	150
WLAN IEEE Standard	802.11b (11 Mbps)
Simulation time (s)	900
Maximum cluster size	50
Ping broadcast interval (s)	1
Variable-value parameters	
Network size (number of nodes)	80; 160; 240; 320; 400
Node maximum speed (m/s)	0 (static); 2 (dynamic)
DiLoC specific parameters	
Number of anchor nodes	25
Anchor node expiry time (s)	50
Anchor node expiry amount	10
Number of APs	25
AP Broadcast TTL	5
C constant used in BC metric	2

Given the enormous quantity of different possible scenarios that the combination of parameters provides, only the most significant were chosen. In particular, the parameters that most influence the scalability of the network are the network size (number of nodes) and the maximum speed that nodes can achieve. Considering the vast application that clustering can have and that this simulation study aims to evaluate a generic scenario, a specific node mobility pattern, like Group Mobility, Freeway or Manhattan models would not be suitable [39]. Thus, a random model, the Random Waypoint, was preferred and a 50 second pause time was chosen. The C value was chosen according to an empirical study to assess the required relevance of the available cluster positions. The chosen value was 2, providing slightly more importance to the amount of available positions in a cluster, in order to embrace the variation of the network size. Each simulation execution was repeated 30 times, assigning to each a distinct seed value.

4.1.1. Special Requirements

In addition to the main simulation parameters, DiLoC has special requirements for proper evaluation. As described before, the NI approach requires anchor nodes, which must be configured before network deployment. The BI approach also requires APs to be present in the scenario. The number of anchor nodes and APs strongly influences the initial network topology, since clusters will begin to be formed around them. A high number of anchor nodes and APs will translate in a high amount of clusters, resulting in fragmented topologies. On the other hand, a low number of anchor nodes and APs will result in clusters with a large number of nodes, which may suffer from scalability issues. Concerning this trade-off, it is necessary to carefully choose an amount of anchor nodes and APs, suitable with the scenario size and amount of nodes, simultaneously ensuring full scenario coverage. To evaluate the impact of location information, all scenarios will have 25 anchor nodes, placed in a grid fashion. A lower amount of anchor nodes would result in less clusters and lower initial overhead. However, this amount of anchor nodes was chosen since it is the minimum required amount to provide full coverage for this scenario size. To be noted that anchor nodes do not represent an extra number of nodes in the network. In fact, for a scenario with a network size of 160 nodes, 25 are initially configured as anchor nodes and remaining 135 as regular nodes. Furthermore, anchor nodes become regular nodes after 50 seconds or when the amount of nodes that joined the cluster is greater than 10.

Typically, lower amount of APs results in less radio interferences, however for an even comparison, the number of APs present for the BI approach is also 25. APs are also placed in a grid fashion, similarly to the initial position of anchor nodes in the NI approach. Finally, the third approach requires both anchor nodes and APs to be present. For this evaluation, 25 anchor nodes and 25 APs, placed in a grid fashion, were also deployed. Once again, the anchor nodes do not represent an additional number of nodes in the network.

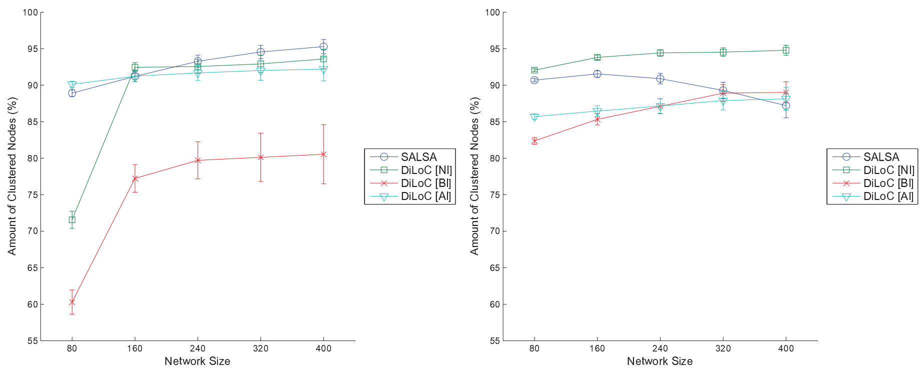
4.2. Results

This section presents the obtained results from the simulation. The evaluation features three main metrics, which will be described further, along with the discussion of results.

4.2.1. Number of Clustered Nodes

This metric provides the number of nodes that are associated with the cluster structure. Nodes that are isolated, i.e. not in-range with any node, cannot be assigned to a cluster. Therefore, since the area of the scenario is constant for all network sizes, there is a bigger percentage or nodes that are likely to be unclustered in smaller networks.

Figure 4 shows the percentage of the average amount of clustered nodes for static and dynamic scenarios. In the static scenario there is a significant increase of clustered nodes from network sizes of 80 to 160. In the 80 nodes scenario,



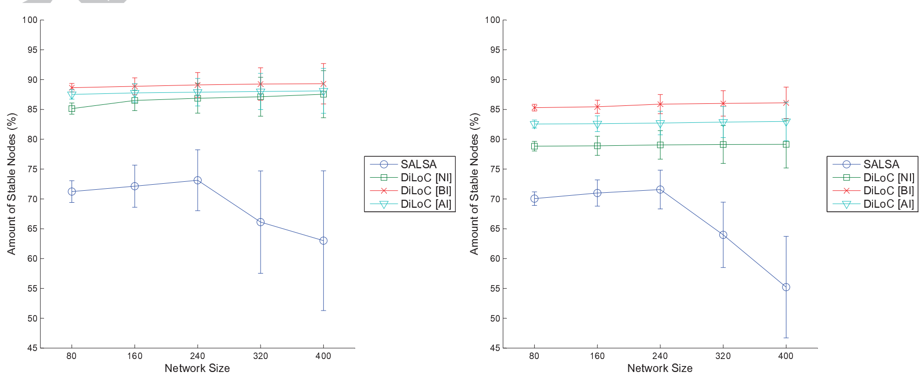
(a) Static (0 m/s) (b) Dynamic (2 m/s)

Figure 4: Average Amount of Clustered Nodes (in percentage)

the connectivity is reduced, due to the low density of nodes. In the 160 nodes scenario, the density is higher, which significantly increases the connectivity and, thus the amount of clustered nodes. Also in the static scenario, SALSA shows good results, particularly for largest networks, however it shows a clear scalability problem in the dynamic scenario. The BI approach presents the lower number of clustered nodes for both scenarios. Since clusters are created around APs and cannot move, there is a great percentage of nodes that are isolated (i.e. not in-range to clusters), thus affecting the overall amount of clustered nodes. The NI approach presents the best overall amount of clustered nodes. The AI approach also presents very good results for static scenarios, however the amount of clustered nodes drops around 4% in the dynamic scenario, even tough it presents a good scalability.

4.2.2. Cluster Stability

The stability of clusters can be measured according to the amount of time that nodes belong to a cluster, without suffering re-clustering operations.



(a) Static (0 m/s) (b) Dynamic (2 m/s)

Figure 5: Average Amount of Stable Nodes (in percentage)

For this analysis, a cluster stability metric is utilized, which defines a stability time (ST), from which nodes are considered to be stable (2).

$$ST = k \times \frac{r \times p}{v \times d} \tag{2}$$

where r is the transmission range of nodes, p is the pause time, v the average of node speed (mean value of minimum and maximum speed), d the density of nodes (number of nodes per Km^2) and finally, k represents an arbitrary constant, equal in all simulation executions, enabling the transformation of the ratio to a real execution time.

The stability metric ST provides a mechanism of determining the amount of nodes that were stable during the simulation for a period greater than the ST value. All nodes, including unclustered nodes, are contemplated in the measurement of this metric. Figure 5 shows the percentage of stable nodes for static and dynamic scenarios.

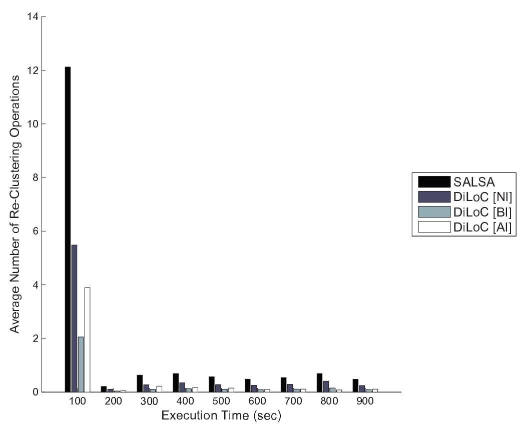
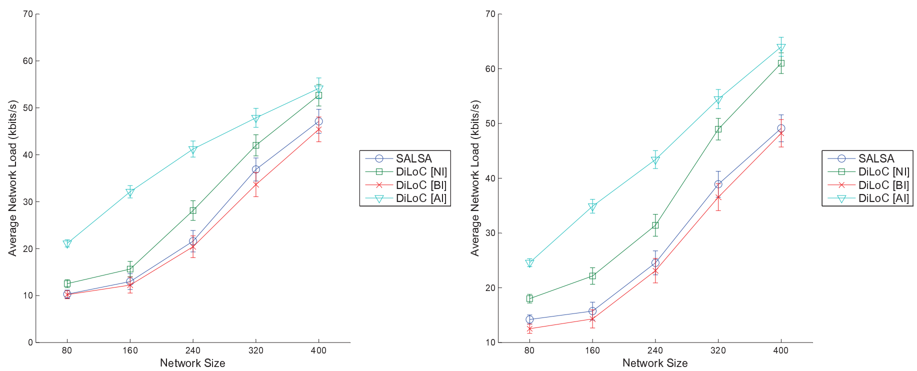


Figure 6: Re-Clustering Operations over Simulation Execution

As a quick first analysis of the Figure, the BI approach clearly outperforms the remaining algorithms. Since nodes are created and maintained around APs, this algorithm can deliver higher amounts of stable nodes. Even in the dynamic scenario, this approach presents similar results to the static scenario in the NI approach. On the other hand, the NI, BI and AI approaches maintain the percentage amount of stable nodes, whereas SALSA begins to severely decrease in networks larger than 240 nodes, presenting high standard deviations due to the large deviation in the number of re-clustering operations (i.e. node assignments to clusters). Generally, DiLoC outperforms SALSA since it always maintains clusters around reference points (either APs or anchor nodes), which leads to a lower amount of re-clustering operations and therefore, a higher stability. To better illustrate this fact, Figure 6 shows the average number of re-clustering operations over execution time, using an equal seed value, for SALSA and DiLoC with AI approach, in a 400 node network with a 2 m/s maximum speed. Overlooking the initial start of the network and the following Random Waypoint pause time (configured to 50 seconds, as shown in Table 4), it can be noticed that SALSA performs a much higher amount of re-clustering operations, when compared to the AI approach. This is mainly due to the random dispositions and shapes of SALSA clusters. Between the three approaches of DiLoC, the NI is the one that presents the lowest amount of stable nodes. Since it does not rely on any infrastructures for cluster maintenance, it maintains clusters with less stability.

4.2.3. Network Load

The network load represents all the received and transmitted traffic in the network. This metric translates the overall weight of the network, including the traffic generated by the MAC layer and the clustering control overhead.



(a) Static (0 m/s) (b) Dynamic (2 m/s)

Figure 7: Average Network Load (in kbits/s)

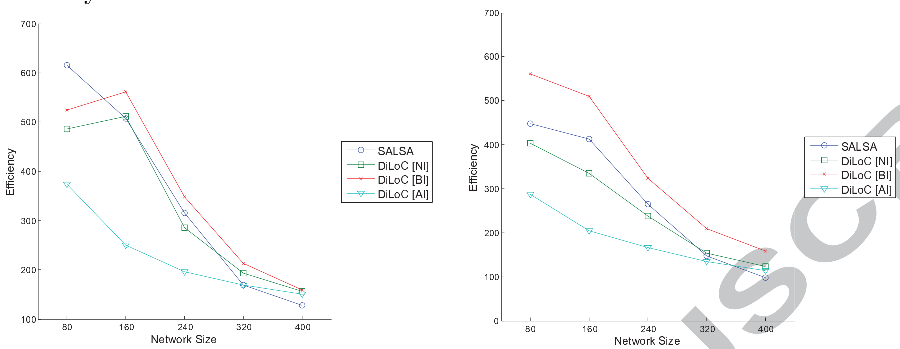
Figure 7 shows the average network load consumed by each approach. As a global overlook, the BI approach and SALSA have a very similar cluster maintenance overhead. The NI approach consumes more bandwidth than SALSA and BI approaches due to the additional communication between anchor nodes and regular nodes, in the initial set up phase. Specifically in this phase, the NI approach consumes an average of around 14.3% of the total network load in the static scenario and 16.8% in the dynamic scenario. During this phase, anchor nodes broadcast a large amount of *Ping* and *Hello* messages, since they are responsible for the creation of clusters and node table updates of the their cluster. Finally, the AI approach has the higher overhead, particularly in the static scenario. The AI approach combines the mechanisms of both NI and BI and additionally it has to decide in which mode to operate, thus consuming a larger amount of traffic. During the initial set up phase, AI consumes a combined average of around 19.1% of the network load in the static scenario and 19.6% in the dynamic scenario.

4.2.4. Efficiency

This metric assesses the overall efficiency of the different approaches. The Efficiency is determined according to equation 3.

$$E = \frac{CN \times SN}{NetLoad} \tag{3}$$

where the CN represents the amount of clustered nodes, SN the amount of stable nodes and $NetLoad$ the network load. Thus, this metric represents a combination of some of the previous evaluated metrics, attempting to represent the efficiency of the clustering schemes. The E value represents the obtained efficiency index value.



(a) Static (0 m/s) (b) Dynamic (2 m/s)

Figure 8: Efficiency (in kbits/s)

Figure 8 shows the obtained efficiency index in static and dynamic scenarios. Since this metric depends on the amount of clustered nodes, it also shows, in static scenarios, an increase of efficiency between the network size of 80 and 160. Despite this fact, SALSA can be more efficient than the remaining schemes for the network size of 80 nodes, in the static scenario. This is due to its high amount of cluster nodes and low network load in such scenario. However, the overall results show that the DiLoC BI approach is the more efficient scheme and the AI approach the less efficient, in both static and dynamic scenarios. This occurs because the NI approach has the lowest network load, and the AI approach has the highest network load. The AI approach does not rely on anchor nodes, taking advantage of the existing infrastructures to provide location information. Thus, the initial overhead caused by anchor nodes does not exist. This concludes that DiLoC is more efficient, when network infrastructures exist. However, such environments may not exist and location information, provided by anchor nodes, may be required.

Summarizing this section, DiLoC presents good performance levels, overcoming SALSA in most evaluation metrics. This shows that the location-awareness mechanism is efficient on providing a stabler network topology with a superior number of clustered nodes. Overall results show that the NI approach is more efficient, thus being the most suitable to scenarios where network infrastructures exist. To infrastructure-less scenarios, both the BI approach and SALSA are suitable, however the BI approach provides a significantly more stable topology, with the trade-off of a slight higher network load. Finally, for hybrid scenarios, the usage of the AI approach is questionable. Despite presenting a higher stability when compared to SALSA, it consumes a significant larger amount of network load. Thus, it may be preferable to utilize the NI approach for such scenarios, disregarding any existent network infrastructures.

5. Conclusion

This work proposed the new DiLoC clustering scheme, based on location information. Three different approaches of DiLoC were studied, creating location extraction mechanisms where infrastructures are not present and also taking advantage of existing network infrastructures.

In the evaluation results the BI approach presents very good levels of stability, however completely fails in the amount of clustered nodes, due to the isolation of nodes. The NI approach is better in the amount of clustered nodes and network load, when compared to AI, however it shows a lower amount of stable nodes. Looking at the overall results, the AI approach presents a good solution to deal with unknown or hybrid scenarios. Even with a lower amount of clustered nodes, it presents very good stability results, however at a higher maintenance overhead. To be noted that the AI approach was simulated with both APs and anchor nodes (distributed in a grid fashion), in contrast with other approaches that were only simulated with the respective one.

Summarizing, DiLoC presents better performance levels in all evaluation metrics, when compared to SALSA. The future of this work contemplates the analysis of the AI approach on more realistic scenarios, regarding remote indoor locations where nodes are potentially not in range with APs.

6. Acknowledgements

This work was supported by a PhD Scholarship (SFRH / BD / 81829 / 2011) from the Portuguese National Foundation for Science and Technology (FCT), the MORFEU project (PTDC / EEA-CRO / 108348 / 2008), the iCIS project (CENTRO-07-ST24-FEDER-002003), co-nanced by QREN, in the scope of the Mais Centro Program, and User Centric Routing (UCR) project (PTDC/EEATEL/103637/2008), also from FCT. The authors would like to thank the OPNET University Program for the licenses provided for the OPNET Modeler Wireless Suite R.

References

References

[1] J. Yu, P. Chong, A survey of clustering schemes for mobile ad hoc networks, Communications Surveys & Tutorials, IEEE 7 (1) (2005) 32–48.

[2] F. Tolba, D. Magoni, P. Lorenz, A stable clustering algorithm for highly mobile ad hoc networks, Systems and Networks Communications, 2007. ICSNC 2007. Second International Conference on (2007) 11–11doi:10.1109/ICSNC.2007.9.

[3] S. Muthuramalingam, M. Sujatha, R. Surya, and R. Rajaram, "An enhanced sectorized clustering scheme based on transmission range for manets," in *Recent Trends in Information Technology (ICRTIT), 2011 International Conference on*, june 2011, pp. 269 –274.

[4] A. Roy, M. Hazarika, and M. Debbarma, "Energy efficient cluster based routing in manet," in *Communication, Information Computing Technology (ICCICT), 2012 International Conference on*, oct. 2012, pp. 1 –5.

[5] J. Tenhunen, V. Typpo, M. Jurvansuu, Stability-based multi-hop clustering protocol, Personal, Indoor and Mobile Radio Communications, 2005. PIMRC 2005. IEEE 16th International Symposium on 2 (2005) 958–962 Vol. 2. doi:10.1109/PIMRC.2005.1651583.

[6] G. Angione, P. Bellavista, A. Corradi, E. Magistretti, A k-hop clustering protocol for dense mobile ad-hoc networks, Distributed Computing Systems Workshops, International Conference on 0 (2006) 10. doi:http://doi.ieeecomputersociety.org/10.1109/ICDCSW.2006.7.

[7] W. Choi, M. Woo, A distributed weighted clustering algorithm for mobile ad hoc networks, Telecommunications, 2006. AICT-ICIW '06. International Conference on Internet and Web Applications and Services/Advanced International Conference on (2006) 73–73doi:10.1109/AICT-ICIW.2006.11.

[8] R. Zoican, An enhanced performance clustering algorithm for manet, in: MELECON 2010 - 2010 15th IEEE Mediterranean Electrotechnical Conference, 2010, pp. 1269 –1272. doi:10.1109/MELCON.2010.5475942.

[9] K. T. Mai, H. Choo, Connectivity-based clustering scheme for mobile ad hoc networks, Research, Innovation and Vision for the Future, 2008. RIVF 2008. IEEE International Conference on (2008) 191–197doi:10.1109/RIVF.2008.4586354.

[10] D. Wei, H. Chan, E. Chuwa, B. Majugo, Mobility-sensitive clustering algorithm to balance power consumption for mobile ad hoc networks, Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on (2007) 1645–1648doi:10.1109/WICOM.2007.414.

[11] Z. Qiang, Z. Ying, G. Zheng-hu, A trust-related and energy-concerned distributed manet clustering design, in: Intelligent System and Knowledge Engineering, 2008. ISKE 2008. 3rd International Conference on, Vol. 1, 2008, pp. 146 –151. doi:10.1109/ISKE.2008.4730915.

[12] C. Huang, Y. Zhang, X. Jia, W. Shi, Y. Cheng, H. Zhou, An on-demand clustering mechanism for hierarchical routing protocol in ad hoc networks, Wireless Communications, Networking and Mobile Computing, 2006. WiCOM 2006.International Conference on (2006) 1–6doi:10.1109/WiCOM.2006.217.

[13] C.-H. Hsu, K.-T. Feng, On-demand routing-based clustering protocol for mobile ad hoc networks, Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. IEEE 18th International Symposium on (2007) 1–5doi:10.1109/PIMRC.2007.4394482.

[14] A. Dana, A. Yadegari, A. Salahi, S. Faramehr, H. Khosravi, A new scheme for on-demand group mobility clustering in mobile ad hoc networks, Advanced Communication Technology, 2008. ICACT 2008. 10th International Conference on 2 (2008) 1370–1375. doi:10.1109/ICACT.2008.4494019.

[15] A. Hussein, S. Yousef, S. Al-Khayatt, O. Arabeyyat, An efficient weighted distributed clustering algorithm for mobile ad hoc networks, in: Computer

Engineering and Systems (ICCES), 2010 International Conference on, 2010, pp. 221 –228. doi:10.1109/ICCES.2010.5674857.

[16] L. Conceição, D. Palma, M. Curado, A novel stable and low-maintenance clustering scheme, in: Proceedings of the 2010 ACM Symposium on Applied Computing, SAC '10, ACM, New York, NY, USA, 2010, pp. 699–705. doi:http://doi.acm.org/10.1145/1774088.1774232.
URL http://doi.acm.org/10.1145/1774088.1774232

[17] R. Want, A. Hopper, V. Falcão, J. Gibbons, The active badge location system, ACM Trans. Inf. Syst. 10 (1992) 91–102. doi:http://doi.acm.org/10.1145/128756.128759.
URL http://doi.acm.org/10.1145/128756.128759

[18] P. Bahl, V. Padmanabhan, Radar: an in-building rf-based user location and tracking system, in: INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, Vol. 2, 2000, pp. 775 –784 vol.2. doi:10.1109/INFCOM.2000.832252.

[19] T. Pfeifer, D. Elias, Commercial hybrid ir/rf local positioning system.

[20] L. Ni, Y. Liu, Y. C. Lau, A. Patil, Landmarc: indoor location sensing using active rfid, in: Pervasive Computing and Communications, 2003. (PerCom 2003). Proceedings of the First IEEE International Conference on, 2003, pp. 407 – 415. doi:10.1109/PERCOM.2003.1192765.

[21] Y.-M. Cheng, Using zigbee and room-based location technology to constructing an indoor location-based service platform, in: Intelligent Information Hiding and Multimedia Signal Processing, 2009. IIH-MSP '09. Fifth International Conference on, 2009, pp. 803 –806. doi:10.1109/IIH-MSP.2009.106.

[22] A. Harter, A. Hopper, P. Steggles, A. Ward, P. Webster, The anatomy of a context-aware application, Wirel. Netw. 8 (2002) 187–197. doi:http://dx.doi.org/10.1023/A:1013767926256.
URL http://dx.doi.org/10.1023/A:1013767926256

[23] A. Smith, H. Balakrishnan, M. Goraczko, N. Priyantha, Tracking moving devices with the cricket location system, in: In Proceedings of the 2nd ACM International Conference on Mobile Systems, Applications, and Services (MobiSys04, ACM Press, 2004, pp. 190–202.

[24] O. Cruz, E. Ramos, M. Ramirez, 3d indoor location and navigation system based on bluetooth, in: Electrical Communications and Computers (CONIELECOMP), 2011 21st International Conference on, 2011, pp. 271 –277. doi:10.1109/CONIELECOMP.2011.5749373.

[25] M. Rodriguez, J. P. Pece, C. J. Escudero, In-building location using bluetooth, in: in In Proceedings of the International Workshop on Wireless Ad Hoc Networks, 2005.

[26] A. Raghavan, H. Ananthapadmanaban, M. Sivamurugan, B. Ravindran, Accurate mobile robot localization in indoor environments using bluetooth, in: Robotics and Automation (ICRA), 2010 IEEE International Conference on, 2010, pp. 4391 –4396. doi:10.1109/ROBOT.2010.5509232.

[27] D. Sanchez, M. Quintana, J. Navarro, Wlan location determination using probability distributions with search area reduction via trilateration, in: Wireless and Mobile Communications, 2009. ICWMC '09. Fifth International Conference on, 2009, pp. 328 –333. doi:10.1109/ICWMC.2009.61.

[28] N. Hassan, S. Hossain, N. Wahab, S. Ariffin, N. Fisal, L. Latiff, M. Abbas, C. K. Neng, An indoor 3d location tracking system using rssi, in: Signal-Image Technology and Internet-Based Systems (SITIS), 2010 Sixth International Conference on, 2010, pp. 323 –328. doi:10.1109/SITIS.2010.59.

[29] P. K. Enge, The global positioning system: Signals, measurements, and performance, International Journal of Wireless Information Networks 1 (1994) 83–105, 10.1007/BF02106512.
URL `http://dx.doi.org/10.1007/BF02106512`

[30] M. Guillemette, I. Fontaine, C. Caron, Hybrid rfid-gps real-time location system for human resources: Development, impacts and perspectives, in: Hawaii International Conference on System Sciences, Proceedings of the 41st Annual, 2008, p. 406. doi:10.1109/HICSS.2008.195.

[31] S. Seidel, T. Rappaport, 914 mhz path loss prediction models for indoor wireless communications in multifloored buildings, Antennas and Propagation, IEEE Transactions on 40 (2) (1992) 207 –217. doi:10.1109/8.127405.

[32] E.-E.-L. Lau, W.-Y. Chung, Enhanced rssi-based real-time user location tracking system for indoor and outdoor environments, in: Proceedings of the 2007 International Conference on Convergence Information Technology, ICCIT '07, IEEE Computer Society, Washington, DC, USA, 2007, pp. 1213–1218. doi:http://dx.doi.org/10.1109/ICCIT.2007.199.
URL `http://dx.doi.org/10.1109/ICCIT.2007.199`

[33] Y. Jiangsheng, Method of k-nearest neighbors.

[34] C.-S. Cheng, H. Chang, Y.-T. Chen, T. Lin, P. Chen, C. Huang, H. Yuan, W. Chu, Accurate location tracking based on active rfid for health and safety monitoring, in: Bioinformatics and Biomedical Engineering , 2009. ICBBE 2009. 3rd International Conference on, 2009, pp. 1 –4. doi:10.1109/ICBBE.2009.5162266.

[35] L. Bras, M. Oliveira, N. Borges de Carvalho, P. Pinho, Low power location protocol based on zigbee wireless sensor networks, in: Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on, 2010, pp. 1 –7. doi:10.1109/IPIN.2010.5647869.

[36] T. Ohta, S. Inoue, and Y. Kakuda, "An adaptive multihop clustering scheme for highly mobile ad hoc networks," in *Autonomous Decentralized Systems, 2003. ISADS 2003. The Sixth International Symposium on*, april 2003, pp. 293 – 300.

[37] L. Conceio, M. Curado, Smart and balanced clustering for manets, in: H. Frey, X. Li, S. Ruehrup (Eds.), Ad-hoc, Mobile, and Wireless Networks, Vol. 6811 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 2011, pp. 234–247, 10.1007/978-3-642-22450-8_18.
URL `http://dx.doi.org/10.1007/978-3-642-22450-8_18`

[38] OPNET, Opnet simulator, http://www.opnet.com/ (1986).
URL `http://www.opnet.com/`

[39] B. Divecha, A. Abraham, C. Grosan, S. Sanyal, Impact of node mobility on manet routing protocols models (2007).

Luís Conceição is a PhD Student at the Department of Informatics Engineering at the University of Coimbra. His research focuses on Wireless Ad Hoc Networks, being currently focused on developing Clustering Schemes for such environments. He was involved in several national and international projects, namelly MORFEU (Multi-Objective Robot Fleet for improved commUnication), EuQoS (End-to-End Quality of Service over Heterogeneous Networks) and WEIRD (WiMAX Extension to Isolated Research Data networks). He also has several conference publications in the area of Clustering mechanisms for Wireless Ad Hoc Networks.