

Healthcare Supported by Data Mule Networks in Remote Communities of the Amazon Region

Mauro MargalhoCoutinho¹, Alon Efrat², Thienne Johnson², Andrea Richa³, Mengxue Liu³

¹CCET, Universidade da Amazônia, Belém, Brazil, margalho@unama.br

²Dept. of Computer Science, University of Arizona, Tucson, USA, {alon, thienne}@cs.arizona.edu

³Computer Science and Engineering, CIDSE, Arizona State University, Tempe, USA, {andrea.richa, mengxue.liu}@asu.edu

Abstract— This paper investigates the feasibility of using boats as Data Mule nodes, carrying medical ultrasound videos from remote and isolated communities in the Amazon region in Brazil, to the main city of that area. The videos will be used by physicians to perform remote analysis and follow up routine of prenatal examinations of pregnant women. Two open source simulators (The ONE and NS-2) were used to evaluate the results obtained utilizing a CoDPON (Continuous Displacement Plan Oriented Network) network. The simulations took into account the connection times between the network nodes (boats) and the number of nodes on each boat route.

Keywords- Telemedicine; healthcare; data mule network; epidemic routing; opportunistic routing; Amazon.

I - INTRODUCTION

Data Mule is an acronym for (Mobile Ubiquitous LAN Extension). It is related to vehicles that physically carry a computer with a storage device and a limited telecommunication module (usually Wi-Fi) between remote areas in order to effectively create a data communication link [1]. A *Delay-Tolerant Network (DTN)* is an architecture that provides a common method for interconnecting heterogeneous gateways or proxies that employ store-and-forward message routing to overcome communication disruptions [2]. Both technologies are tolerant to disconnections in the network and are often complementary. However, despite the fact that these technologies do not adhere to the principle of ubiquitous computing, they open the doors to hundreds of applications whose integration was not possible earlier, mostly due to the high costs, or even infeasibility, of implementing a networked infrastructure in some scenarios.

One of these scenarios is the Marajó Archipelago, located in the Amazon region of Brazil, which occupies a large fluvial area of 104,142 Km² (approximately 40,700 sq mi) [24]. The largest island on the archipelago, the Marajo Island, has roughly the same size as Switzerland and is the largest island in the world to be surrounded by freshwater. According to the last Brazilian census [3], only 43% of the archipelago's population of 487,010 inhabitants lives in urban areas. There are no bridges linking the islands to the mainland, and boats are by far the main mean of transportation in the region. Some cities, like Afua (known as the "Venice of the Amazon") are completely built on water. In these places, socio-infrastructure problems abound, especially health related problems as there are few physicians available in the remote communities. Most

medical care is done through sporadic government programs that involve the displacement of a team to serve the population, in particular in the outlying areas. The goal of this paper is to present an alternative telemedicine infrastructure, focused on performing remote triage and prioritization of medical care in these remote regions of the Amazon rain forest. We implement a telemedicine infrastructure over a CoDPON¹ (Continuous Displacement Plan Oriented Network) data mule system [4], commonly applied in scenarios where nodes have pre-defined routes. We present a preliminary evaluation of the capabilities of our proposed telemedicine infrastructure by means of simulations. We simulate the implementation of a remote ultrasound video analysis service for pregnant women in the Marajó archipelago, as part of their prenatal treatment, over our data mule network.

This paper is organized as follows: Section II presents related work, Section III presents the system architecture, Section IV presents a description of the remote ultrasound telemedicine application, and Section V presents our simulation studies. Section VI concludes this work and presents suggestions for the improvement of services through the use of fountain codes.

II - RELATED WORK

A. DTN and Data Mule

The first terrestrial DTN applications were proposed for sensor networks [5][6][7]. Soon other applications emerged for integration of remote communities like Daknet [8] that provided low-cost digital communication, allowing remote villages to leap from the expense of traditional connectivity solutions and the deployment of a full-coverage broadband wireless infrastructure. Another proposal, named MotoPost, was presented in [9]. The concept was developed to address the communication and information access needs of remote rural villages in India that do not have access to communication technologies.

The Epidemic Routing is a routing protocol for DTN. This algorithm is named this way because messages are spread like infectious diseases [20]: when two nodes come into contact, the first node sends the second node a summary vector containing information that uniquely identifies the bundles of

¹ Research is partly supported by NSF grants CCF-1116368 and CCF-1353089 and FIDES (In Brazil)

information (messages) it has. The second node then sends the first node any bundles it has that the first does not. This takes place in both directions whenever two nodes come into contact [21]. For a moderate volume of data, epidemic routing is considered as one of the most effective routing schemes for delay-tolerant networks; indeed, epidemic routing allows for shortest routing delays [22].

These and other efforts sensitized the scientific community to produce new tools to evaluate the impact of DTNs in different scenarios. One of the most popular DTN simulators, "The ONE" [7] -- Opportunistic Network Environment -- has rich features available for simulating DTNs with numerous mobility models. It is a useful tool for platform for implementation and evaluations of new ideas. Recently, Osman Khalid *et al* [10] evaluated the concept of Checkpoint (CP). CPs are autonomous high-end wireless devices with a large buffer for storage and are responsible for temporarily storing messages to be forwarded. According to the authors, the basic components of a CP node are processor, memory, solar powered batteries, multiple interfaces (Standard Ethernet, 802.11b/g/n, Bluetooth), GPS, and storage. The results of simulation in an area of 4 x 3 KM (the size of the city of Fargo, ND, USA), show that message delivery ratio improves by increasing the number of checkpoints in the area.

B. "Store and forward" telemedicine

In scenarios where real-time communication network and might not be a practical solution for the delivery of large files, "store and forward" approaches could come handy. Typically these systems encapsulate and then transmit medical information to the recipient, without accepting immediate acknowledgment. These methods might be required due to cost constrains. Examples of "store-and-forward" include correspondence via e-mail, fax or Data Mule. The main advantage of this form of telemedicine is that no online services are required and the content of the information can be examined as soon as possible [11]. The use of Delay Tolerance Network infrastructure to provide telemedicine in remote areas has also been investigated in [12], and a simple telemedicine system using a digital camera was evaluated in [13]. In that test-bed, radiographs on a viewing box were photographed at a remote hospital in South Africa, using a digital camera, and the resultant images were stored in JPEG format and transmitted, later to be read on a PC monitor by radiologists in Durban and Cape Town.

III - CODPON ARCHITECTURE

CoDPON network, proposed in [4], is a specific Data Mule system inspired by the air traffic control system: same "modus operandi", where boats operate in a similar way to airplanes, with pre-defined routes. It basically consists of four components: Nodes, DACTs (Data Application in Transit - data divided into logical units), Displacement Plans, and the Communication Protocol. There are three types of nodes: Vehicles, Peers Base Stations (PBS) and Hot Spots. *Vehicles* (in our case- boats) act as mules, carrying the data spread

along the path. The vehicles used on a CoDPON network operate on predefined routes, with pre-scheduled (days and times) departures and arrivals. All boats have a computer housed in an airtight box and use solid state disks to ensure more robust and persistent storage. The solid state is to prevent damage due to motion of boats.

Peer Base Stations are fixed nodes, previously mapped, where vehicles make scheduled stops. Peer Base Stations are usually located on the river shore. Hot Spots are special nodes that exist only in areas that have connection to the Internet. They work as a faster gateway between CoDPON nodes. Unfortunately, solutions such as the ones proposed in [1], which reduce latencies equipping the MULEs with an always-on connection (such as a cellular or satellite phone), are not available in some scenarios (for example, there is a very limited telecommunications coverage in Amazonia). A DACT represents a minimum unit of data transferred between nodes and has a self-meaning that allows pieces of data to be separated from others without losing consistency.

A *displacement plan* is applied to all boats. It contains basic information about the entire route of the boat, including starting point, destination, stopovers and estimated time of journey and anchoring. *Journey time* is the time required for a boat to travel its entire route. *Anchor time* usually is the time for passengers to embark or disembark at village piers. Each boat has its own displacement plan stored on board, including a table containing the hydrographic distances between peer base stations.

The communication protocol is based on the principle of Data Mule. It does not require changes in the TCP stack, as is the case with the DTN.

IV - APPLICATION

A major difficulty for health care in isolated areas, like the Marajó archipelago, occurs due to the small number of physicians who reside in the region. The vast majority of those just make periodic visits to the communities in the region. Among the most affected by the infrequent medical visits in the communities are pregnant women, who need regular monitoring. Hence, we considered the proposed application of remote ultrasound monitoring. The CoDPON system will act as a data mule network, carrying and delivering ultrasound files from the remote communities (Fig. 1) to the state capital, Belem, for analysis by physicians (Fig. 2).

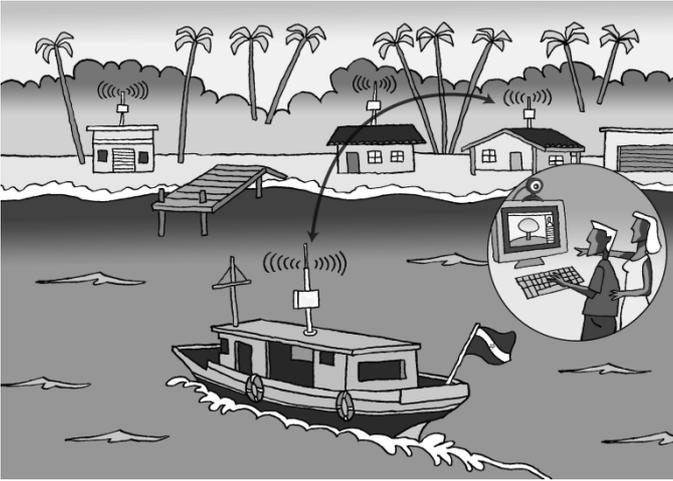


Fig. 1 – Boats exchanging ultrasound files

The ultrasound parameters used in the simulation process were obtained by consulting a team of experts on ultrasounds, including OB\Gyn doctors in Belem, who operate in the Marajó region. According to the team, the ultrasound examination is a dynamic diagnostic method in which the physicians acquire the ultrasound videos and interpret them in a dynamic way. Two synchronized videos are required for the remote analysis to be efficient: the first video shows the image generated by the equipment and the second video shows the examination process (the position of the transducer being used by health agent on the patient’s body).

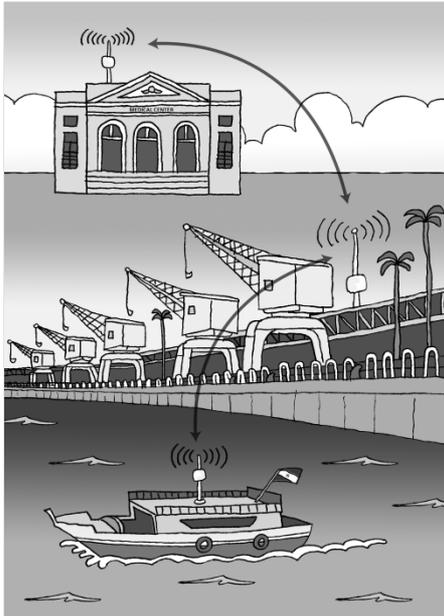


Fig. 2 – Delivery of ultrasound files at the State capital.

V - SIMULATION

Two open source simulation tools were selected for the evaluation of the scenario described in the Marajó Archipelago. The first software was NS-2, probably the most

widely used simulator in the scientific community [14]. The NS-2 wireless module was used to inspect the connection time between boats (in motion) and Peer Base Stations and between a boat and another boat (both in motion) (Fig.3). The other software was The ONE simulator [7], which has rich features available for simulating DTNs with numerous mobility models.

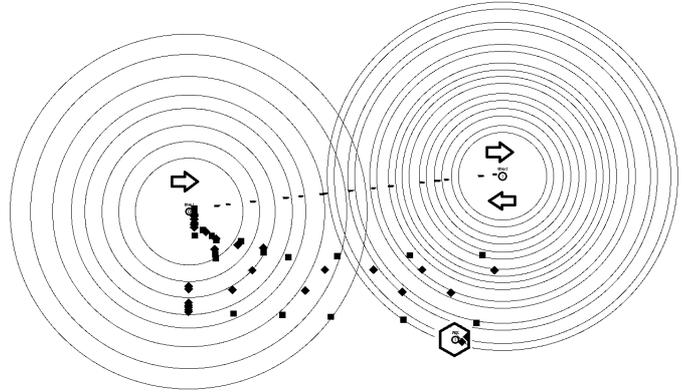


Fig. 3 – Connection time Simulation

A. Scenario

For the simulation, an area of 104,142 Km² (approx. 40,700 sq mi) comprising the Marajó Archipelago, located in the state of Pará, Brazil was selected (the south part of the Marajó Island is shown in Fig. 4). The routes for the boats were created according to local data, obtained from the office of the public boats terminal in the city of Belém (capital of Pará). The boats are equipped with Wi-Fi IEEE 802.11 devices and they follow routes throughout the simulation area, following predefined routes. Each boat has scheduled stopovers at various Peer Base Stations.

B. Simulation Parameters

For the simulation, various combinations of parameters were selected (Table 1).

TABLE 1 - SIMULATION PARAMETERS

Parameter	NS-2	The ONE
Flat Grid / World Size (Km)	2000x200	22000x20000
Simulation Time	1000sec	216000 (60 hours)
Wireless Interface Transmit Speed (Mbps)	11	11
Wireless Interface Range (m)	100	100
Radio Frequency	2.4 GHz	-
Path Loss Exponent (Rain)	2	-
Average boat speed (Km/h)	15	15
Number of PBS (DTN Module in Seashore)	1	11
Displacement Plans	-	9

The scale of the simulations conducted in NS-2 and The ONE simulator were very different: The simulation in NS-2,

we evaluated the amount of data transmitted between boats and PBSs, which required a finer grain simulation and more specifics about the radiowave communications. The simulations conducted in The ONE simulator concerned the flow of packets over the network, following the end-to-end boat displacement plans.

As some routes are traveled for up to seventeen hours, the Time to Live (TTL, maximum time that a message could still live before arrive to the destination) value was set to 1,122 minutes, 10% more than the value of the maximum travel time. When the TTL expires the message will be self-destroyed, if it had not been delivered yet.

C. Traffic Generator

In order to configure the traffic generator for the simulator, the demand of real data loads in Marajó were investigated *in loco*. During the sporadic government health programs, physicians perform one hundred ultrasound exams in one weekend, once a month. In other words, working at maximum capacity, they perform fifty exams per weekend day. If we consider that each ultrasound size, in Mpeg format, is around 5 Megabytes (using the minimum acceptable standards for digital compression of a fetal ultrasound video-clip [16]), and we add the video of a health agent performing the examination, 10 Megabytes is a reasonable size for each ultrasound video set. Thus, the traffic generator of the ONE simulator was set up to work at 384kbps. The value was obtained considering that fifty ultrasounds per day are performed, each generating a 10 MB ultrasound file. Three PBSs, located at the extreme points of the Marajó archipelago, were selected as sources of video transmissions, all of them sending files to the main PBS in Belém. The average boat speed of 15 Km/h was obtained by GPS *in loco*.

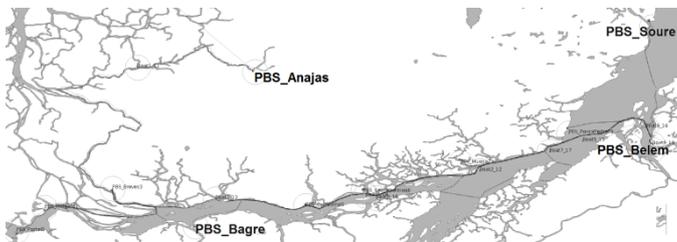


Fig. 4 – Area overview

VI - RESULTS

A. Simulation in NS2

Once a boat stops at a PBS, it they could receiving DACTs, due to its large buffer (SSD disk). Therefore, the investigations were concentrated in critical points, that is, when the data transfer occurs while the boats are moving. In the simulation, three different scenarios were tested.

Scenario 1 shows a boat passing within 50 meters of another boat in the opposite direction. In this case, both boats are in motion. Scenario 2 shows a boat passing within 50 meters of a Peer Base Station, which is a fixed point, without stopping. Scenario 3 shows two boats in movement, moving in

the same direction, one traveling at twice the speed of the other. In all scenarios, connection times and amount of data transferred were analyzed (Fig 5).

In the first scenario, transmission occurs between seconds 172 and 320 of the simulation time line, which represents a duration of 148 seconds. This results in a total of 2.3 minutes of connection time between the boats. This would allow for the transfer of two ultrasound video sets.

Number of Megabytes Transferred Between Nodes

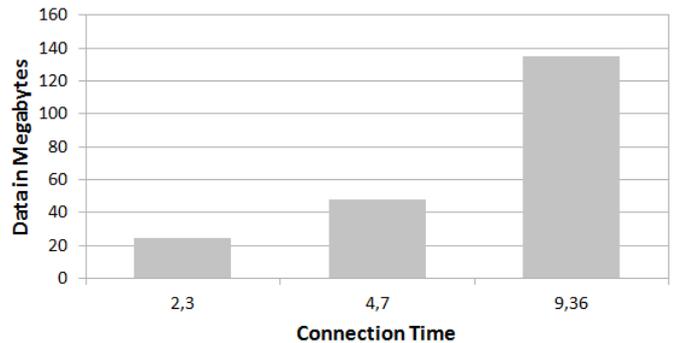


Fig. 5 – Amount of data transferred

In the second scenario, transmission occurs between seconds 99 and 381 of the time line simulation, which represents a total of 282 seconds. This results in a total of 4.7 minutes of connection time between a boat and a PBS. With these values it would be possible to transfer 47.62 Megabytes, which would allow the transfer of four ultrasound video sets.

In the third scenario, transmission occurs between instants 198 and 760 of the time line simulation, which represents a total of 562 seconds. This results in a total of 9.36 minutes of connection time between the boats. With these values, 135.08 Megabytes would be transferred, which would allow the transmission of thirteen ultrasound exams sets (Fig. 6).

Number of Ultrasounds Transferred Between Nodes

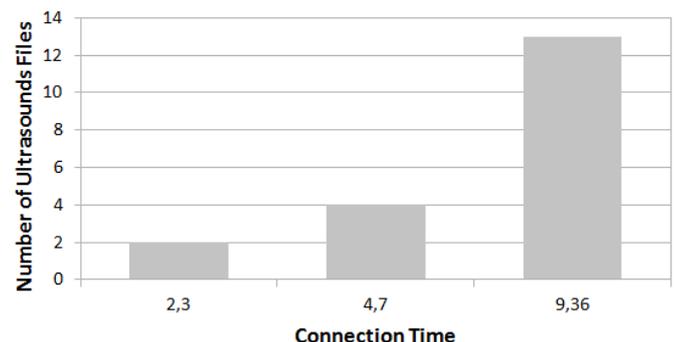


Fig. 6 – Number of medical exams transmitted

B. Simulation in ONE

The map of the Marajó archipelago was obtained from

IDESP (Institute of Economic, Social and Environmental Development of Pará) [15]. Using the open source GIS tool OpenJump. The map was post-processed and marked with various locations such as villages, small cities, and the route of the boat lines.

The scenario used in this section aims to transfer ultrasound sets from remote communities of the Amazon to Belém. The ONE simulator scenario comprises 11 PBS, and only three of them (PBS_Anajás, PBS_Bagre, and PBS_Soure) generated ultrasound exam files. Each boat has an associated route file defined by the OpenJump software, based on real routes obtained from the main boat station in Belém.

Epidemic routing was used in the simulations. The evaluation was performed by changing the number of boats in the entire scenario, not only between source and destination. The metrics were compared with one, three, five and seven boats acting in each route.

Figure 7 shows the relation between started, dropped, relayed and aborted packets. *Started packets* are the packet transmissions started between network nodes. *Dropped packets* are the packets that are dropped from nodes' buffers and relayed packets are the packets successfully transmitted between nodes. As it can be seen, the number of dropped packets rises with the number of boats, due to the epidemic routing, which indicates a rapid saturation of the network capacity. Basically this is due to having more boats carrying packets the "wrong" way (in the opposite direction of the PBS it should be going to), which end up "wandering" in the network without ever making it to Belém.

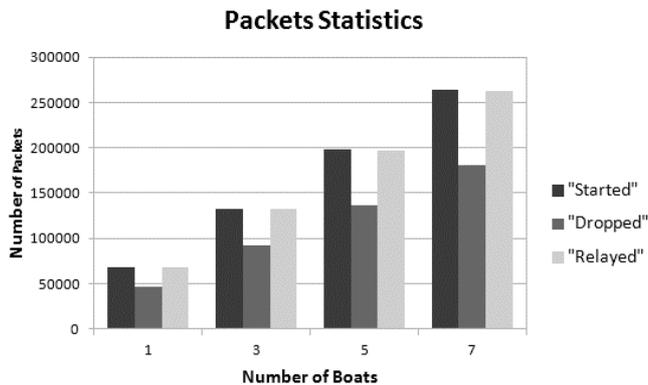


Fig. 7 – Packet Statistics

Despite the fact that the delivery probability increases with the increased number of boats [Fig. 8], the amount of wandering packets in the network grows in a much larger proportion [Fig. 9], also rapidly saturating the network capacity. This is clear when the overhead ratio is observed. For an epidemic-like protocol, the overhead ratio can be interpreted as the number of replicas for each message transmission. In an environment without QoS, this could be a problem because of the large amount of messages to be delivered.

Delivery Probability per Number of Boats

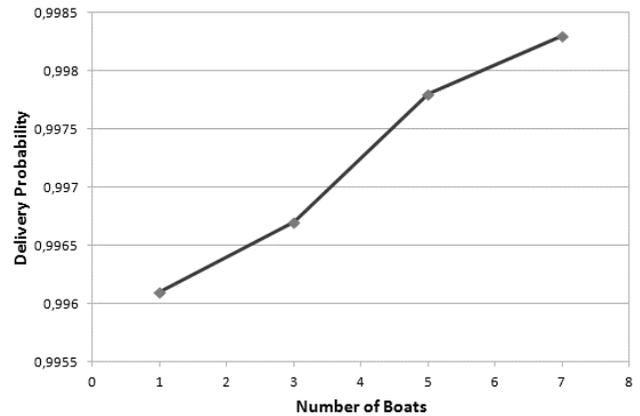


Fig. 8 – Number of Megabytes

Without a prioritization mechanism, the most important messages can be delivered later and may not be relayed in time. In a CoDPON network, QoS prioritization is based on the Olympic service (QoS architecture that classifies messages in gold, silver and bronze: gold goes first, then silver, and bronze is considered best effort) [17] to avoid this problem.

Figure 8 shows the delivery probability, which is the probability related to the number of successfully delivered messages. This calculation excludes the messages sent in the opposite direction with boats not traveling to the destination. It grows linearly accordingly to the increase in the number of boats. This information can be used by network administrators and local government to estimate how efficient the service is. Depending on the results, more boats can receive the necessary equipment to become nodes in a CoDPON network. Figure 10 shows the average message latency (in seconds) from creation to delivery. It drops as more boats are used to carry the packets.

The simulation performed can be used to scale the number of boats that will act as data mules in the system and the impact of the use of traditional routing protocols like epidemic routing, for example, to decrease the latency with an increased number of boats.

Overhead Ratio per Number of Boats

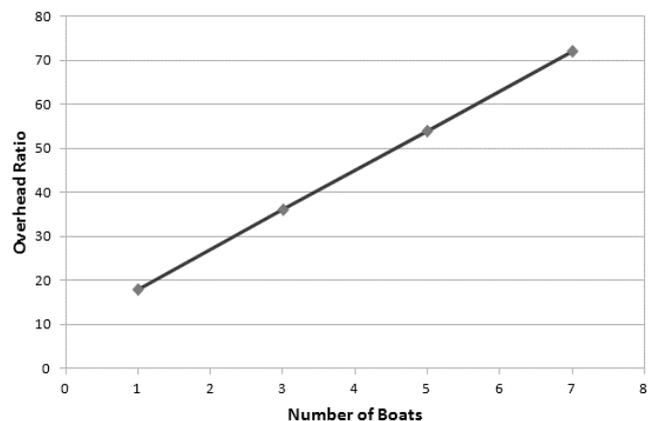


Fig. 9 – Overhead Ratio

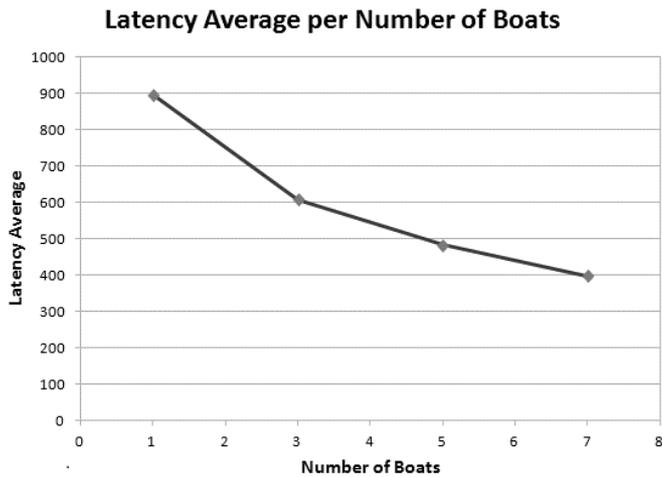


Fig. 10 – Latency Average (in Seconds)

It is clear that the proposed system has a high delivery ratio with high overhead. But as our goal is to provide medical services to underserved communities, the delivery of the messages is our priority, thus the overhead is the issue that needs to be improved but it is acceptable as a starting point.

CONCLUSIONS

The demand for health services is large in the Amazon region. However, due to the lack of communication and transportation infrastructure, many patients who require special care end up unattended. One of the most important aspects of our proposed system refers to the inclusion of health experts in the patients' medical files preliminary analysis. For example, if a large number of neurological problems is detected in the analysis of the remote preliminary examinations files, neurologists may be included in the group of doctors in the boat caravan to the villages.

The simulations in this paper clearly indicate that epidemic routing creates a very large burden on the network, since the packet spreads in all directions without discretion. In a network with known routes, as is the case of CoDPON, where the PBSs may act similarly to CheckPoint [10] spreading packets, the spread can be controlled to minimize the amount of unnecessary transfers, since there is some control available through the displacement plans.

Future Work

For future work, we are developing an algorithm that will diffuse messages according to the displacement plan of the boats in a controlled way. That algorithm will be implemented for the ONE simulator and the results will be compared with the existing implementation.

Another topic for further research is to evaluate how to reduce excessive messages sent via the epidemic routing protocol. The Spray and Wait [19] protocol limits the number of copied messages to a predefined maximum and distributes the copies to the contacted nodes until the maximum number of copies has been reached. Supp-Tran [23] optimizes the performance of epidemic routing protocol in terms of delivery

probability, overhead ratio and Hop count average. Such approaches and other message distribution protocols will be evaluated to improve our system's performance.

Since components of the proposed system might suffer unpredictable failures, robustness must be introduced. In order to increase efficiency and robustness in the delivery of exam files, we will study the use of Fountain Codes. According to [18], a Fountain codifier is a metaphoric well which produces an infinite quantity of drops of water (packages of codified data). Thus, anyone who wishes to receive a coded file puts a bucket under the well and collects drops until the number of drops in the bucket is sufficient to recover the original file and decode the packages. In the analogy of the CoDPON system, the well would be the Checkpoints, and the drops (packages) would be transported by the boats up to the destination (the state capital), which would have a bucket to collect them. If one assumes that the information one wishes to be transferred is a 'bucketful of water', one needs only to generate several drops and feed the Bucket (which will be collected at the destination) at the checkpoints, until they are full of original information. The great contribution of Fountain Codes is that they can retrieve files, even without all the original packages, or it is possible to rebuild a lost package using copies left over from packages already received.

REFERENCES

- [1] Shah, Rahul C., et al. "Data mules: Modeling and analysis of a three-tier architecture for sparse sensor networks." *Ad Hoc Networks* 1.2 (2003): 215-233.
- [2] RFC 4838, Delay-Tolerant Networking Architecture. <http://tools.ietf.org/html/rfc4838>.
- [3] IBGE Censo 2010. <http://www.censo2010.ibge.gov.br/>
- [4] Coutinho, Mauro Margalho, et al. "A new proposal of data mule network focused on Amazon riverine population." *Proc. of the 3rd Extreme Conference on Communication: The Amazon Expedition*. ACM, 2011.
- [5] Sundararaj, Laveen, and Palanisamy Vellaiyan. "An Overview of Alagappa University Delay Tolerant Water Monitoring Network." *IJCSNS* 10.5 (2010): 19.
- [6] Anastasi, Giuseppe, Marco Conti, and Mario Di Francesco. "Data collection in sensor networks with data mules: An integrated simulation analysis." *Computers and Communications, 2008. ISCC 2008. IEEE Symposium on*. IEEE, 2008.
- [7] Keränen, Ari, Jörg Ott, and Teemu Kärkkäinen. "The ONE simulator for DTN protocol evaluation." *Proc. of the 2nd international conference on simulation tools and techniques*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009.
- [8] Pentland, Alex, Richard Fletcher, and Amir Hasson. "Daknet: Rethinking connectivity in developing nations." *Computer* 37.1 (2004): 78-83.
- [9] Shrikant Naidu, Suresh Chintada, Moushumi Sen and Seshadri Raghavan "Challenges in deploying a delay tolerant network." *Proc. of the third ACM workshop on Challenged networks*. ACM, 2008.
- [10] Khalid Osman, Khan Samee, Kolodziej Joanna, Zhang Limin, Li Juan, Havat Khizar, Madani Sajjad, Wang Lizhe, Chen Dan. "A Checkpoint Based Message Forwarding Approach For Opportunistic Communication." *ECMS*. 2012.
- [11] Smith, Anthony C., et al. "Telemedicine and rural health care applications." *Journal of postgraduate medicine* 51.4 (2005): 286.

- [12] Pandian. Poondi Srinivasan. et al. "Store and forward applications in telemedicine for wireless IP based networks." *Journal of Networks* 2.6 (2007): 58-65.
- [13] Corr. Peter. et al. "A simple telemedicine svstem using a digital camera." *Journal of Telemedicine and Telecare* 6.4 (2000): 233-236.
- [14] Kalvanpur. Ariun. et al. "Implementation of an International Teleradiology Staffing Model 1." *Radiology* 232.2 (2004): 415-419.
- [15] The Network Simulator ns-2. Available at: <<http://www.isi.edu/nsnam/ns>>. 09/11/2012.
- [16] IDESP (Institute of Economic, Social and Environmental Development of Pará). <http://www.idesp.pa.gov.br/>
- [17] Begg. L., et al. "Minimum acceptable standards for digital compression of a fetal ultrasound video-clip." *Journal of telemedicine and telecare* 7.suppl 2 (2001): 88-90.
- [18] MacKav. David JC. "Fountain codes." *IEE Proceedings-Communications* 152.6 (2005): 1062-1068.
- [19] Spvropoulos. Thrasvoulos. Konstantinos Psounis. and Cauligi S. Raghavendra. "Spray and wait: an efficient routing scheme for intermittently connected mobile networks." *Proc.of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*. ACM, 2005.
- [20] Le. Viet-Duc. Hans Scholten. and Paul Havinga. "Evaluation of opportunistic routing algorithms on opportunistic mobile sensor networks with infrastructure assistance." *International Journal On Advances in Networks and Services* 5.3 and 4 (2012): 279-290.
- [21] Ivan Christian, Christian H. W. Oey, and Sangman Moh, "A Simple Temporal Node Mobility Model for Performance Analysis in Delay Tolerant Networks", *JCIS: Journal of Communications and Information Sciences*, Vol. 3, No. 3, pp. 200-206, 2013.
- [22] Wang. Oingshan. and Zygmunt J. Haas. "Analytical model of epidemic routing for delay-tolerant networks." *Proc. of the 1st ACM workshop on High performance mobile opportunistic systems*. ACM, 2012.
- [23] Elkouch, Rachid, Mostata Bellafkih, and Nabil Ababou. "New Strategy to Optimize the Performance of Epidemic Routing Protocol." *International Journal of Computer Applications* 92 (2014).
- [24] Map of Maraió Archipelago. (2014). <https://www.google.com.br/maps/place/Arquipélago+do+Marajó,+PA/@-0.9950443,-50.0457297,9z/data=!4m2!3m1!1s0x9298b193c0b8372d:0x5ab a6a2a87297515>