Utiliser la densité des utilisateurs mobiles dans les grandes villes afin de délivrer des SMS

Yannick Léo, Anthony Busson, Carlos Sarraute, Eric Fleury

Utiliser la densité des utilisateurs mobiles dans les grandes villes afin de délivrer des SMS

Yannick Leo[14] et Anthony Busson[14] et Carlos Sarraute[2] et Eric Fleury[134]

[1] *ENS de Lyon, Lyon, France*
[2] *Grandata Labs, Buenos Aires, Argentina*
[3] *INRIA, France*
[4] *University of Lyon (UMR CNRS - ENS Lyon - UCB Lyon 1 - INRIA 5668)*

Dans cet article nous proposons un protocole de réseau semblable à un DTN afin d'acheminer les SMS en se basant sur une étude de mesures et une analyse des SMS provenant d'une trace sur une période de deux mois. Plus précisément nous réalisons une analyse spatiale et temporelle du réseau cellulaire de Mexico en prenant en compte les messages géo-localisés. L'analyse temporelle nous permet de détecter les évènements et de vérifier les périodes de surcharge comprenant un trafic anormal ou inattendu et d'étudier l'évolution de paramètres classiques tels que l'activité ou la distance entre l'expéditeur et le receveur. L'analyse spatiale est basée sur le diagramme Voronoï des stations de base couvrant la ville de Mexico. Nous expliquons comment le trafic de SMS peut être caractérisé. Une telle caractérisation clef nous permet de répondre à la question suivante : est-il possible de transmettre des SMS en utilisant les téléphones comme des relais dans une grande ville comme Mexico ? Nous avons défini un protocole de réseaux simple afin de transmettre des SMS d'un point de départ à un point d'arrivée. Ce protocole semblable à un DTN ne nécessite pas de routage ni de connaissance globale. Le protocole profite de la localité des SMS, de la densité de téléphone à Mexico et de la mobilité des utilisateurs. Nous avons étudié une base de données de téléphones portables incluant 8 millions d'utilisateurs vivant à Mexico. Cela nous a donné une estimation précise du temps de transmission moyen et de la performance globale de notre approche. Après 30 minutes, la moitié des messages ont atteint leur destination avec succès.

Mots-clefs : Réseaux informatiques et télécommunications; Protocoles des réseaux;Systèmes de performance;Science des réseaux;DTN;Protocole de routage

1 Introduction

The need of communicating in a dense city is always increasing. Every day, millions of SMS are sent in a large city like Mexico City. Whereas, traditional SMS is challenged by alternative messaging services, SMS is still a growing market and remains a very popular service over cellular networks since 82% of mobile users are sending SMS in our dataset. The SMS is well-known and well-used in both developed and developing countries, and uses this communication standard with penetration in more than 220 countries.

During rush hours or special events, SMS traffic may consume a non negligible part of the backbone network capacity, and sometimes saturates it. It is becoming a great challenge to increase the amount of traffic delivered to the users while keeping the infrastructure stable (i.e., same number of relays and backbone capacity).

One goal is to demonstrate that the DTN approach could be feasible and helpful. We clearly do not target an implementation of our approach in existing 3GPP standards or existing cellular networks protocols. We demonstrate by replaying millions of real SMS traffic the reliability and gain of our proposals. Note that such approach could be used in future cellular standards, and also be used as a key enabler for P2P applications.

The contributions of this study are two-fold. We perform an analysis in time and space of SMS sent during two months in Mexico city [SBB14] including event detection (skip in this paper). The second contribution is the proposal of two protocols that aim to carry SMS traffic, and to relieve the infrastructure network in terms of load. We use the same real SMS traces to evaluate the efficiency of these less centralized protocols.

2 Data source

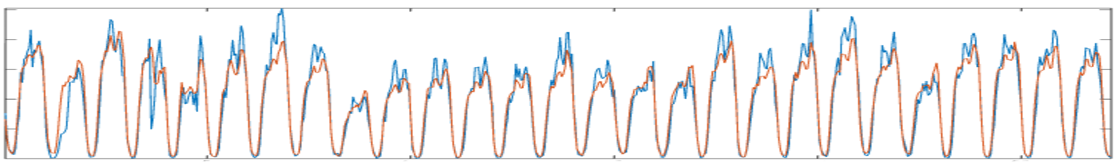

FIGURE 1: Plot of the global SMS activity in Mexico over 2 months (in blue) and the sampling one based on SMS for which the location of the source and the destination is known (in red).

We used traces extracted from mobile phones [SBB14] representing 92 millions of clients in Mexico and 8 millions in Mexico City during a 2-month period : from March to April 2014. The city is covered by 775 base stations that are part of the telecommunication network of a cellular operator. This anonymized dataset contains 70 million SMS and 170 million phone calls all over Mexico. Some calls are localized, *i.e.,* we know the base station of the source or destination. From these mobile calls, it has been possible to localize in Mexico City 1.5 millions of SMS for our study. For each SMS, we set the localization of the closest call in time. We assume that if the source and the destination received or made a localized call 30 minutes before or after the SMS, we can effectively set a source and destination location for the SMS. We also noticed that more than 92% of geolocalized SMS sent from Mexico City have the destination in Mexico City.

For our study, we define a graph with the base stations as nodes, and wherein there is a link/edge between two nodes if they are neighbors in the Voronoï tessellation : there is a link between two base stations (bs_1, bs_2) if and only if bs_1 and bs_2 have a common border in the Voronoï tessellation built according to the base station locations. Consequently, we can define a distance between two base station not only in term of kilometers but also in terms of hops through this graph. Two neighbors are at distance 1, and the neighbors of the neighbors are at a distance 2, and so on. As we do not have any hierarchical information on base stations, we consider that every base station has the same role.

3 Protocols

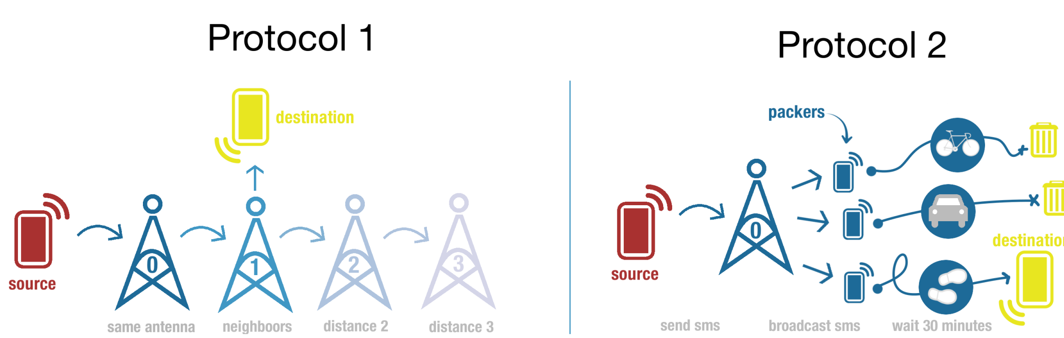

FIGURE 2: Sketches of the protocols 1 and 2.

We describe two protocols 2 to deliver the SMS in another way than the protocol used in the cellular network. Our approach relies on the density and mobility of phone users combined with the locality of SMS. The protocols use the base stations that are close to the base station attached to the source and users that are connected to it.

Protocol 1 The source sends the SMS in the usual way to the base station to which it is attached (Figure 2). Then, this base station retransmits to the stations around itself which are neighbors (all the stations that are at a distance 1). This process can be repeated in order to reach the neighbors of the neighbors at distance 2 from the original station. Considering a fixed distance k, the SMS is well received if and only if the destination is attached to a base station at a distance $\leq k$. The efficiency of Protocol 1 depends on the locality of the exchanges. During this process, the location of the destination does not have to be known.

Protocol 2 The second protocol relies on the mobiles attached to the original base station of the sender (Figure 2). When this base station receives a SMS, it duplicates this message to the mobiles which are attached to that station. In practice, every mobile in that cell gets the SMS, they are called the packers. If the destination is one of these packers, the SMS is already transmitted. The packers who have just received the SMS move through Mexico City and may switch from that antenna to others. As the packers are moving, if a packer and the destination are at the same time attached to the same base station, the destination will receive the SMS. At this moment, the packer sends the SMS to the base station and the base station sends the SMS to the destination as it is attached to this antenna. The success of this communication depends on the density of packers and thus of mobile users and on their mobility. If many packers are moving randomly all over the city, the probability of reaching the destination has to be very high.

4 Results

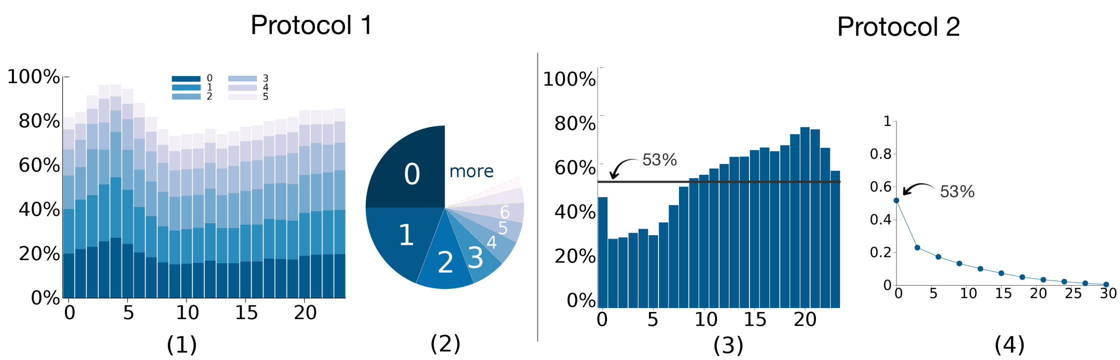

FIGURE 3: (1) Temporal transmission success for Protocol 1 as function of the hours of the day. We consider different number of hops for the transmission of SMS ; (2) global transmission success according to the number of hops averaged over the day. (3) Temporal transmission success for Protocol 2 averaged by day hours where packers keep the message 30 minutes then remove it from their phones ; (4) inverse cumulative distribution of the global transmission success according to the delay ($\leq$ 30 min).

Results for Protocol 1 We quantified the efficiency of the first protocol in Figure 3 for which the success is only based on the locality of SMS. We show that 20% of SMS sent reached their destination if we only consider people attached to the same base station ($k = 0$), almost 35 % when the base station transmits the SMS to its neighbors ($k = 1$) and 46 % if we consider base stations at a two-hop distance ($k = 2$). This result highlights the fact that the destination is usually close to the base station of the source. In comparison, if we consider the base station network of Mexico City, the average of hops between two random stations is 11 and the diameter (maximum distance in hops) of this network is 22. Considering $k \leq 5$, only 13% of the base stations are reachable. In that sense, we can say that there is a high locality of SMS.

Results for Protocol 2 In the traces that we studied for the network of Mexico City (1.5 million SMS over 2 months), one third of SMS are delivered in less than 10 minutes and half of them are transmitted in 30 minutes (Figure 3), this delay is set arbitrarily and reasonably. We consider that after 30 minutes without being delivered, the transmission fails. One can notice that 20% of transmission success is due to a packer that switch from an antenna to another. The mobility is a key point for that protocol. When the activity is high, during rush hours, the protocol has more than 70% of delivering success. It is interesting to notice that the efficiency of our protocol is the best when the operator service is challenged. Moreover, as we miss out some locations (only 9% of the SMS were localized), we just have a part of the packers and have an underestimation of the efficiency. The protocol 1 relies on the locality of SMS sent during the day whereas the protocol 2 also takes advantage of the mobility and density of phone users leading to a transmission success rate that increases drastically during the rush hours with a maximum rate of 70%.

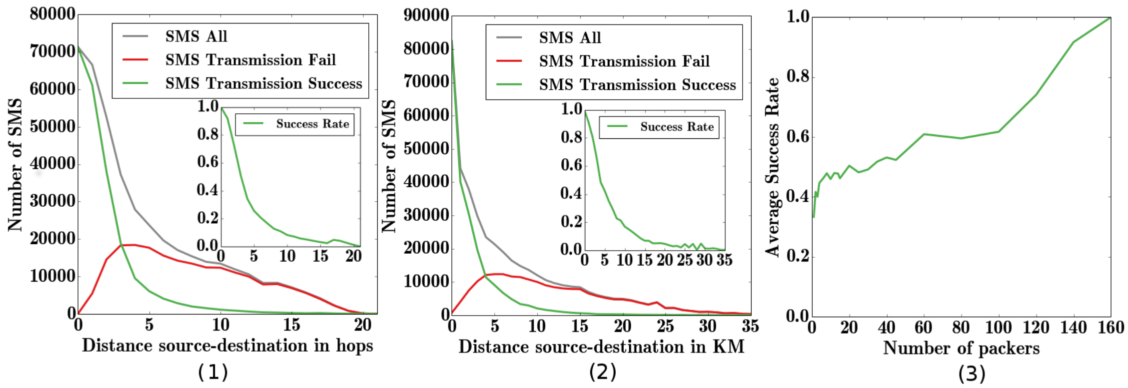

FIGURE 4: Analysis of the success rate of the protocol 2 according to the original distance between the source and the destination. (1) Number of SMS that reach the destination, number of SMS that are dropped and total number of SMS according to the distance in hops of the SMS ; (1, focus) Success rate according to the distance in hops of SMS ; (2) Number of SMS that reach the destination, number of SMS that are dropped and total number of SMS according to the original distance in km of the SMS ; (2, focus) Success rate according to the original distance in km. There is the original distance between the source and the destination when the SMS is sent and the delivering distance that corresponds to the distance between the source and the destination when the message is delivered ; (3) The average success rate, percentage of SMS delivered, according to the number of packers.

The distance in hops or in km between the source and the destination has a great impact on the success rate. If the source is far from the destination, the packers should be very quick and lucky to reach the destination within 30 minutes. As we can notice on the figure 4 (focus), the success rate is decreasing drastically as the original distance between the source and the destination is increasing. The packers have a key role too as they deliver SMS by moving along the large city. In our dataset, we only have 6% percent of them. In the figure 4, we show that the success rate is increasing as the number of packers is bigger. This figure not only shows that our protocol success rate is actually an underestimation but it gives us an undirect measure about the mobility of phone users. Like the density of users and the locality of interactions, the mobility is a capital asset for the mobile user ad hoc network.

5 Conclusions

In this paper, we have proposed two simple protocols based on DTN principles to carry a part of the SMS traffic in a cellular network. They have been evaluated through an original method as we had the opportunity to replay a large trace of geolocalized SMS in Mexico City. This evaluation has shown that the density of the network, users mobility and locality of SMS increases the efficiency of the protocols and so could unload the backbone network of the operator. Our protocols help to diversify the way of communicating.

Références

[SBB14] C. Sarraute, P. Blanc, and J. Burroni. A study of age and gender seen through mobile phone usage patterns in mexico. In *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on*, pages 836–843, Aug 2014.