Use of Unsupervised Learning Clustering Algorithm to Reduce Collisions and Delay within LoRa System for Dense Applications

Mohammed Alenezi
School of Electronic Engineering and Computer Science
Queen Mary University of London
London, UK
m.alenezi@qmul.ac.uk

Kok Keong Chai
School of Electronic Engineering and Computer Science
Queen Mary University of London
London, UK
michael.chai@qmul.ac.uk

Shihab Jimaa
Department of Electrical and Computer Engineering
Khalifa University
Abu Dhabi, UAE
shihab.jimaa@ku.ac.ae

Yue Chen
School of Electronic Engineering and Computer Science
Queen Mary University of London
London, UK
yue.chen@qmul.ac.uk

***Abstract*—Internet of Things (IoT) is one of the most cited terms within the wireless communication research communities. Next generation wireless networks technologies are expected to have massive-connections of tens of billions of devices. In terms of wireless networks, and in regards to collisions and transmission delay drawbacks being critical challenges when deploying IoT devices, Low Power Wide Area Networks (LPWAN) technologies are considered to be a potential solution for IoT applications. In particular, this paper investigates the use of Long-Range (LoRa) technology for serving dense applications. Furthermore, it identifies a dense application and investigates the possibility of using LoRaWAN for such applications. This work proposes a priority scheduling technique based on unsupervised learning clustering algorithm (K-Means). The proposed technique shows a reduction of the collision rate, the transmission delay and enhancement of the throughput in comparison to conventional LoRaWAN networks and other optimisation techniques.**

***Index Terms*—IoT, LoRa, Clustering, Collision Rate, Throughput, Transmission Delay**

I. Introduction

The world is anticipating hundreds of billions [1], [2] of low battery-powered devices (sensors). Such sensors are to be used across a range of applications. These applications include but not limited to, smart cities, smart agriculture, smart homes and buildings, smart supply chain and logistics, smart industrial control, smart metering and smart street lightening. Devices that transmit and receive smaller amounts of data have been referred to as Internet of Things (IoT). A number of protocols have been developed for Low Power Wide Area Networks (LPWAN). LPWAN protocols include NarrowBand-IoT (NB-IoT), SigFox and LoRa. LoRa is gaining popularity over NB-IoT and SigFox due to support by the industry such as LoRa Alliance, IBM and more. LoRa provides bitrate from 100 to 10000 bit/s, it provides long range coverage ($> 10km$) and has simple and effective MAC. LoRa has less complex modulation and in return less power consumption, mobility support and provides separate operation style (can be integrated within any network as ”do it yourself style”) [3].

LoRa protocol Long Range Wide Area Networks (LoRaWAN) provides high immunity to interferences. This is due to the adoption of Chirp Spread Spectrum (CSS) technique within LoRa physical layer. For this reason, it has been the centre of attention for research communities and manufactories as a potential technology for serving IoT devices. Medical [4], agriculture monitoring [5] and smart city [6] applications are well known areas where LoRaWAN was deployed.

LoRa provides three different device classes; Class A, devices transmit following ALOHA protocol which is the lowest in power consumption; Class B, devices listen periodically for beacons from the gateway; Class C, devices listen almost continuously which is the highest power consumption class.

In [7], LoRaWAN scalability is assessed based on a number of given devices connected to the network. They showed packet-loss of 90% due to collisions in pure ALOHA. Results from [7] were used for this work simulation results validation. In [8], a decoding algorithm that enhance the resolve of overlapped signals is proposed. Results from [7] showed that their proposed algorithm outperforms conventional LoRaWAN in terms of delay. However, their work was based on devices listening for beacons from the gateway (LoRa Class B devices). This class is far more power consuming than Class A (ALOHA-type), specifically for applications with energy constrains.

This paper proposes a priority scheduling technique based on K-Means clustering algorithm, unsupervised machine learning algorithm [9], [10]. The main purpose of this technique is to reduce the collisions within LoRa. Based on prioritising the clusters of which has more priority to send over the other, results show a reduction of collision rate down to 56%. This reduction led to enhancing the overall throughput as well as minimising the total transmission delay in a high dense application. The rest of the paper is organised as follows: Section II reveals the system model and the proposed priority scheduling technique. Section III discusses the simulation results and the optimisation of the proposed technique. This paper is then concluded in section IV.

II. System Model and Proposed Priority Scheduling Technique

For this system model, K-Means clustering algorithm is used to partition n number of LoRa nodes into k clusters in which each LoRa node belongs to the cluster with the nearest distance to its centroid. In this system model, a forest application is adopted as an example of LoRa dense application (Fig. 1).

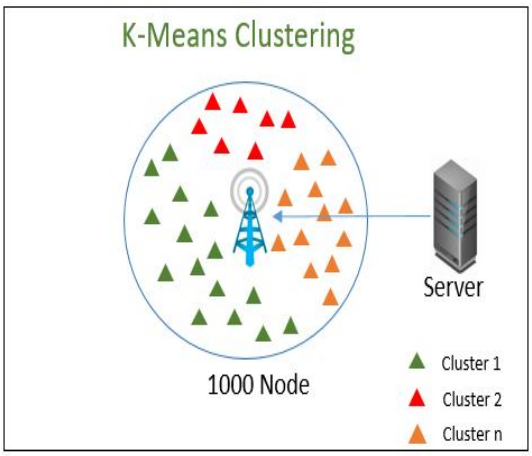

Figure 1. Dense Application using LoRaWAN and K-Means clustering algorithm

In this system model, 1000 LoRa Class A nodes are distributed around one gateway following Poisson random distribution. All nodes are assumed to be transmitting over one channel using one spreading factor (SF7) with a bandwidth of 125kHz and a Coding Rate (CR) of 1 which maximises the actual packet bits on the expense of the Forward Error Correction (FEC) redundant bits.

The nodes fall within different clusters following K-Means clustering algorithm. These clusters were based on set of values $S = \{s_1, \ldots, s_n\}$ where S is in the space of a positive integer number k. For simplicity; these values (A) and (B) are represented by temperature in Celsius and the atmospheric humidity percentage to resemble a forest scenario. Values (A) and (B) are two types of information (metrics) associated with each node. Based on these two values, the node falls within the most suitable level of cluster priority. Considering the 1000 nodes $n = \{n_1, \ldots, n_{1000}\}$, temperature (B) values range from 30° to 45° were assigned following a random-uniform distribution to all the nodes mimicking a summer season. Similar to assigning temperatures, atmospheric humidity (A) values range from 30% to 70% were assigned to the nodes. K-Means is adopted with set of clusters $k = \{3, 4\ and\ 5\}$. Final clusters centres are then prioritised by the network server as in Fig. 2, Fig. 3 and Fig. 4. The purpose is to reduce the collision rate and the transmission delay.

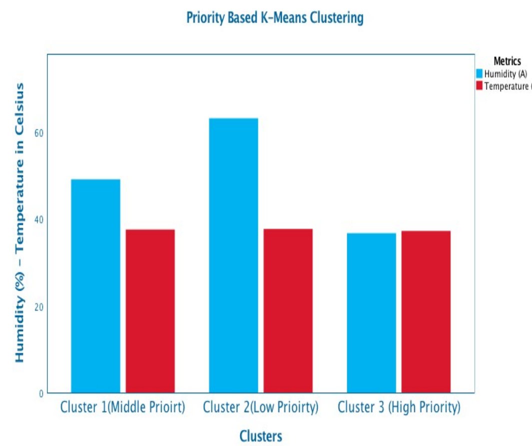

Figure 2. Priority Based K-Means Clustering ($k = 3$) Bar Chart

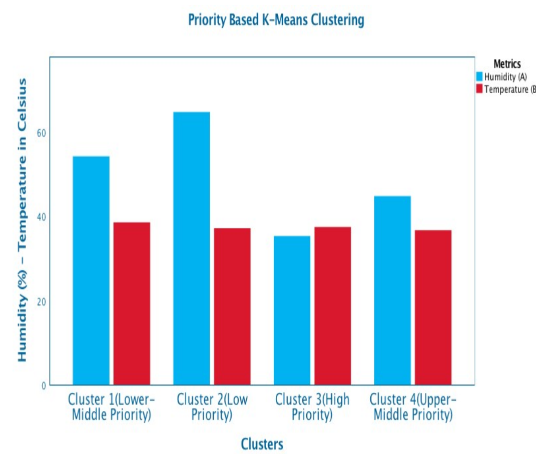

Figure 3. Priority Based K-Means Clustering ($k = 4$) Bar Chart

A. Collision Rate

As defined in the above system model, there are 1000 nodes assumed to be transmitting randomly to the gateway. Each node transmits one (25 bytes) packet following LoRa Class A (ALOHA-type) protocol. A collision happens when two nodes transmit at the same time and frequency. According to Poisson distribution, the probability P of a collision happening is given by (1):

$$P = e^{-2G} \tag{1}$$

Where G is the rate of packet transmission attempts per node. In LoRa typical deployment each node transmits whenever there is a ready to send packet regardless of any other transmission occupying the channel [3]. This behaviour results in a collision rate of up to 91.3% at a LoRa network with 1000

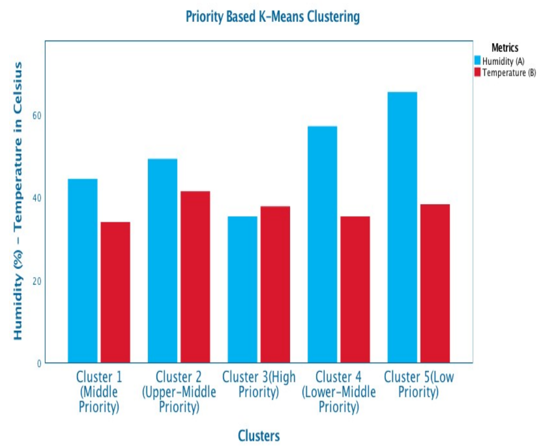

Figure 4. Priority Based K-Means Clustering ($k = 5$) Bar Chart

nodes as shown in Fig. 5. This result is validated with similar collision rate results in [7].

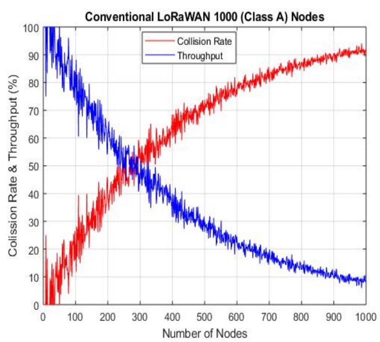

Figure 5. Collision Rate and Throughput in Conventional LoRaWAN (Class A)

B. Transmission Delay

The transmission delay is a function of the number of bits and the bit rate per second [11]. It is proportional to the number of bits within a packet and it is given in (2):

$$Transmission\ Delay = \frac{Numberofbits}{Bitrate} \tag{2}$$

Where Bitrate is given by (3):

$$Bitrate = \frac{SF \times BW}{2^{SF}} \times \frac{4}{4+CR} \tag{3}$$

Where SF is fixed to 7, $BW = 125kHz$ and $CR = 1$. In this system model the packet size is fixed to 25 bytes. Hence using (2) the transmission delay per packet is 0.036 second.

According to LoRa specifications , a collision means retransmission attempt. Hence there is Initial Transmission Delay (ITD) and Retransmission Delay (RTD). The proposed priority scheduling technique introduces waiting times, which means nodes in clusters with lower priority wait until transmissions by nodes in higher priority clusters finish.

C. Proposed Priority Scheduling Technique

The proposed technique is a priority scheduling technique which is based on the aforementioned three, four and five clusters. This technique partitions nodes into clusters based on the priority of transmission. As shown in Fig. 2, Fig. 3 and Fig. 4, two values out of (A) and (B) were set to determine the priority level of each cluster (highest, middle or lowest transmission priority). Nodes in highest transmission priority cluster suppress nodes in other clusters and transmit. Nodes in middle transmission priority clusters transmit only if there are no active transmissions from nodes in higher priority clusters; nodes in lower priority clusters get suppressed. Nodes in lower transmission priority clusters transmit only if there are no active transmissions from nodes in higher clusters. However, this way of prioritising introduces a waiting time which effects the total transmission delay of the whole network. Hence the transmission delay for each cluster is:

$$C_{HP}D = C_{HP}ITD + Colli_{HP}RTD \tag{4}$$

$$C_{MP}D = C_{HP}D + C_{MP}ITD + Colli_{MP}RTD \tag{5}$$

$$C_{LP}D = C_{MP}D + C_{LP}ITD + Colli_{LP}RTD \tag{6}$$

Where $C_{HP}D$ is Higher Priority Cluster Delay, $C_{HP}ITD$ is Higher Priority Cluster Initial Transmission Delay and $Colli_{HP}RTD$ is Higher Priority Cluster Collided packets Retransmission Delay. $C_{MP}D$ is Middle Priority Cluster Delay, $C_{MP}ITD$ is Initial Transmission Delay and $Colli_{MP}RTD$ is Collided packets Retransmission Delay. $C_{LP}D$ is Lower Priority Cluster Delay, $C_{LP}ITD$ is Initial Transmission Delay and $Colli_{LP}RTD$ is Collided packets Retransmission Delay.

From (4), (5) and (6), the Total Transmission Delay (TTD) for LoRaWAN is then calculated as:

$$TTD = C_{MP}D + C_{MP}D + C_{LP}D \tag{7}$$

III. Discussion and Simulation Results

Simulation results (Fig. 6) show both collision rate and the throughput for LoRaWAN using K-Means clustering algorithm. The first row of Fig. 6 shows LoRaWAN performance at $k = 3\ clusters$ the three clusters mean collision rate is reduced to 56% in comparison to conventional LoRaWAN collision rate of 91.3%. As a result the network throughput is further enhanced. At $k = 4\ clusters$ (second row of Fig. 6), the mean collision rate is further reduced to 47%. Furthermore, the collision rate is reduced even more with a mean collision rate of 38% at $k = 5\ clusters$ (third row of Fig. 6). This is due to a lower number of nodes competing on the same channel and transmission time slots. As noticed in Fig. 6, the throughput being a function of successful non-collided packets is enhanced with reducing the number of collisions. Fig. 7 shows total transmission delay comparisons between a LoRa typical network and the proposed priority scheduling technique both serving network with 1000 nodes.

From Fig. 7, it can be shown that typical LoRa transmission delay for serving a network of 1000 nodes reaches up to almost

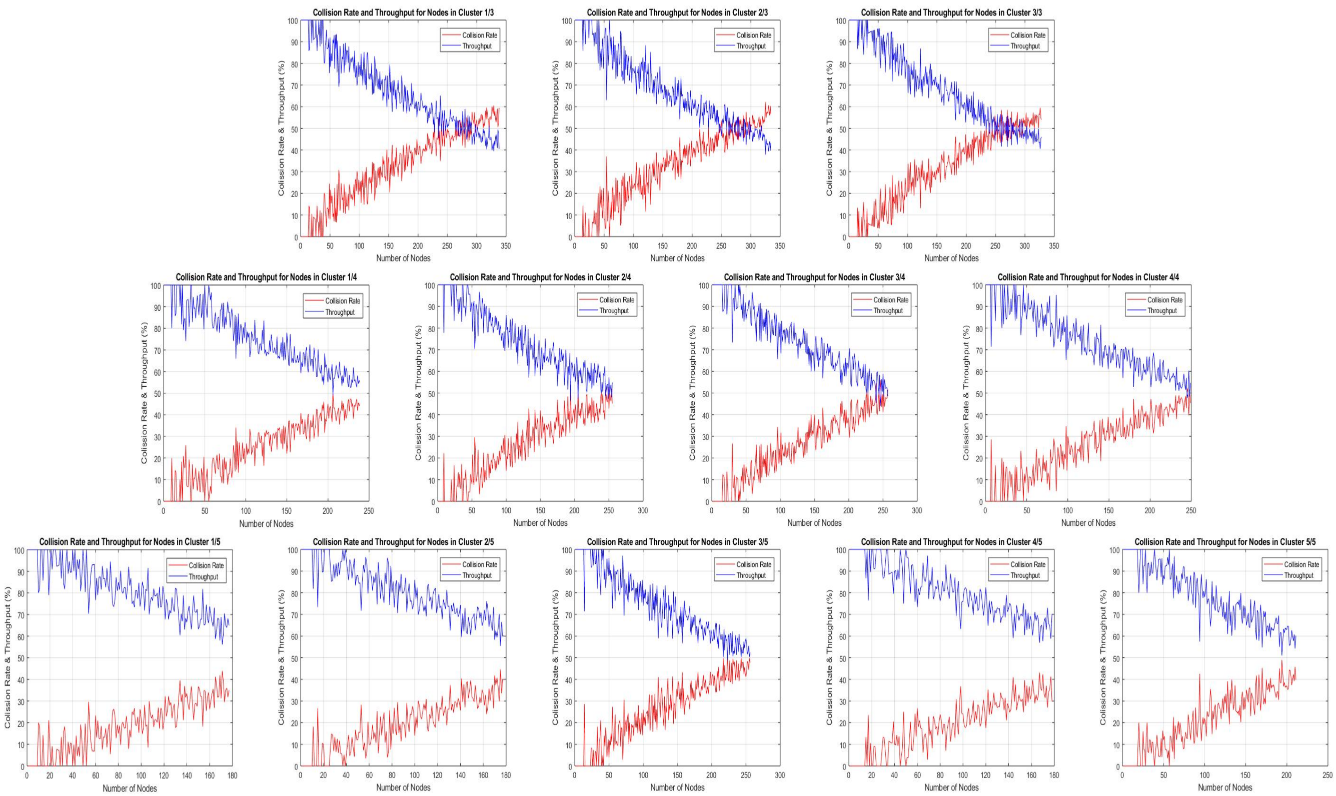

Figure 6. Collision Rate and Throughput for LoRaWAN adopting three clustering scenarios; First Row - $k = 3$ clusters; Second Row - $k = 4$ clusters; Third Row - $k = 5$ clusters

$70s$. This is due to the retransmissions attempts as a result of the high collision rate showed in Fig. 5. In [8], obtained similar results for typical LoRa but for smaller network with up to 300 nodes. The proposed priority scheduling technique for $k = 3$ $clusters$ reduced the delay to $57.6s$. This reduction is a result of reducing the collision rate of the whole network as shown in Fig. 6. In $k = 4$ $clusters$ scenario, the delay is reduced further to $54.1s$. Again this is a result of minimising the collisions as shown in Fig. 6. In $k = 5$ $clusters$ scenario, the delay is even further reduced to $51s$ due to further minimisation of the collisions as shown in Fig. 6.

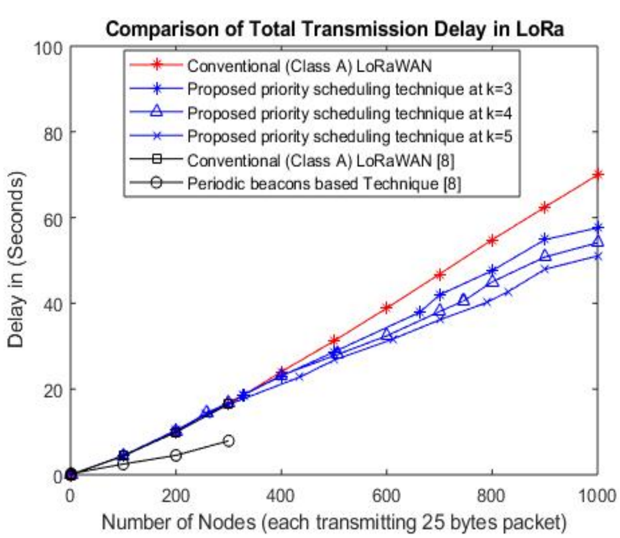

Figure 7. Proposed Priority Scheduling Technique Optimisations of LoRaWAN (Class A) Transmission Delay vs. Conventional LoRaWAN and (Class B) Overlapping Signals Decoding Algorithm [8]

As in (7), collided packets retransmission have severe impact on the total delay. Hence reducing collision rate results in reducing the total transmission delay.

Although it is again for a smaller network, in [8], Periodic Beacons Based Technique shows the lowest delay results. However, their technique is based on periodic beacons which falls within Class B LoRa devices. Class B based devices listens to beacons produced by the gateway and transmit only when there is an available beacon. Apart from the fact that this type of devices is different from the one considered to be used in this work which is LoRa Class A devices (ALOHA-type), Class B is far more power consuming which defeats the whole purpose of using LoRaWAN for such dense applications.

IV. Conclusion

As shown in the simulation results above, priority scheduling technique efficiently reduces the collision rate, the transmission delay and enhances the throughput. Simulations were carried out based on a network with only 1000 nodes. Having more nodes to be served may lead to different results. At this stage of work it is concluded that the use of unsupervised machine learning algorithm (K-Means) to group the nodes into a number of clusters has shown positive enhancements to the overall performance. Having more clusters has shown better performance results. LoRaWAN has a number of pros that makes it a potential technology for handling IoT applications at large scale. Obtained simulation results show that LoRa can potentially form an excellent LPWAN technology to be adopted for the ultra-dense applications. However, unlike the licensed radio frequency which regulates the transmitted signals, the challenge for LoRa nodes is on how to avoid the chaos in the air and reduce the chances of high interferences and hence reduce collisions. Since LoRaWAN operate on license-free frequency bands, some places apply 1% duty cycle limitation on transmission which means each node have limited time to transmit. This limitation if applied would introduce another phase of transmission delay and hence a severe impact on the overall performance.

References

[1] Huawei Technologies, "Building a Trusted and Managed IoT World," Feb 2017. [Online]. Available: https://www.huawei.com/minisite/iot [Accessed May 2019]

[2] Cisco, "Internet of Things," 07 Feb 2019. [Online]. Available: https://www.cisco.com/ [Accessed 29 May 2019]

[3] LoRa Alliance, "LoRaWAN™ Specification v1.1," 11 October 2017. [Online]. Available: https://lora-alliance.org/ [Accessed 14 June 2019]

[4] Philip A. Catherwood, David Steele, Mike Little, Stephen Mccomb, James Mclaughlin, "A Community-Based IoT Personalized Wireless Healthcare Solution Trial," IEEE, Jour. Transl. Eng. Health Med, vol. 6, Basel, Switzerland, May 2018, pp. 1-13.

[5] Davide Sartori, Davide Brunelli, "A smart sensor for precision agriculture powered by microbial fuel cells," IEEE Sensors Applications Symposium (SAS), Catania, Italy, April 2016, pp. 1-6.

[6] Ferran Adelantado, Xavier Vilajosana, Pere Tuset-Peiro, Borja Martinez, Joan Melia-Segui, "Understanding the Limits of LoRaWAN," IEEE Communications Magazine, vol. 55, September 2017, pp. 34-40.

[7] Jetmir Haxhibeqiri, Floris Van den Abeele, Ingrid Moerman, Jeroen Hoebeke, "LoRa Scalability: A Simulation Model Based on Interference Measurements," MDPI, Sensors Journal, May 2017, pp. 1-25.

[8] Nancy El Rachkidy, Alexandre Guitton, Megumi Kaneko, "Collision Resolution Protocol for Delay and Energy Efficient LoRa Networks," IEEE Transactions on Green Communications and Networking, vol. 2, April 2019, pp. 535-551.

[9] J.B. Macqueen, "Some Methods of Classification and Analysis of Multivariate Observations," Proc. Fifth Berkeley Symp. Math. Statistics and Probability, 1967, pp. 281-297.

[10] Anil K. Jain, Richard C. Dubes, "Algorithms for Clustering Data," Prentice Hall, 1988, pp. 89-132.

[11] SEMTECH, "LoRa™ Modulation Basics," May 2015 . [Online]. Available: https://lora-developers.semtech.com/library/product-documents/ [Accessed 24 Apr 2019]