Crowdsourcing Low-Power Wide-Area IoT Networks

Keyi Zhang and Alan Marchiori
Department of Computer Science
Bucknell University, Lewisburg, PA
Email: {kz005, amm042}@bucknell.edu

***Abstract*—The Internet of Things (IoT) promises to allow everyday objects to connect to the Internet and seamlessly interact with users and other machines. For this vital Internet connection, most current IoT devices use a personal gateway device such as a smartphone or a home WiFi access point. The necessity of configuring and maintaining these gateways presents an additional burden for both users and developers of IoT applications. Our vision for IoT connectivity is to eliminate the need for the personal gateway by developing crowdsourced low-power wide area networks (csLPWAN). Recent technologies such as RPMA, LoRa, and R-FDMA enable links to reach 15km using ISM-band transceivers, making csLPWANs an attractive option. In this paper, we investigate the practicality of csLPWANs and develop the first csLPWAN planning tool, PlanIt, which combines topography-aware RF signal analysis with demographic data to predict LPWAN coverage in specific geographic areas. Using PlanIt, we find that most cities achieve 99% coverage by deploying a single LPWAN base station within the city. To provide better service on the csLPWAN, we propose and evaluate DQ-N, a near-optimal MAC protocol to enable efficient bandwidth sharing in highly utilized networks. In the future, csLPWANs could accommodate a heterogeneous set of IoT applications, simplifying the IoT application development cycle, reducing total system cost, improving application reliability, and enhancing the user experience.**

I. Introduction

The Internet of Things (IoT) continues to grow as a new paradigm in which information and communication systems are embedded in our surroundings, enabling new services and applications. This paradigm has been applied in various application domains including intelligent healthcare, environment monitoring, precision agriculture, and smart cities. The IoT ecosystem spans monitoring, storage, communication, and analytical tools. However, creating a scalable and robust communication system for IoT is essential for success [1].

In typical applications, the IoT device stores data locally and then periodically communicates to the cloud via a gateway device. Gateway devices could be smartphones, stationary WiFi access points (AP), or cellular networks. Low-power communication between the IoT device and gateway device using standards such as Bluetooth, ANT+, Zigbee, IEEE 802.15.4, or IEEE 802.11 (WiFi) achieve a maximum communication range typically less than 100 meters. To achieve longer range communication, cellular networks have historically been the only option. This solution is relatively expensive and power hungry as cellular networks were not designed to support IoT devices. Emerging network standards aimed at optimized cellular networks for IoT devices are in various development stages (e.g., LTE-M, NB-IoT, EC-GSM-IoT, etc.), however, these are not yet widely deployed.

Recently, several long-range low-power wireless technologies have been developed including RPMA, LoRaWAN, and R-FDMA (Section II) to address the IoT connectivity problem. These wireless standards claim between 5 km to 15 km communications range using shared spectrum in the ISM bands at low data rates, typically 1 kbaud or less. While still less than the 35 km to 100 km range made possible by higher power cellular radios operating in assigned frequency bands [2], these technologies nevertheless enable the creation of low-power wide-area networks (LPWANs).

In this paper, we propose crowdsourced LPWANs (csLPWANs) as a solution to the IoT connectivity problem and analyze how they would perform in various demographic regions of the United States. We define a csLPWAN as any LPWAN where the base stations are randomly deployed by users of the system rather than deployed in a coordinated fashion by a network operator. As several vendors roll-out private LPWAN networks using fixed infrastructure (e.g., SIGFOX, Ingenu), we demonstrate that csLPWANs are a viable alternative solution for providing IoT device connectivity.

In our preliminary analysis of existing LPWAN protocols (Section II), we discovered that these technologies rely on relatively simple contention-based MAC protocols which have a well-known utility upper bound of 36.8% [3]. Enhancements can be made to increase the channel utility at the cost of increasing energy consumption, latency, and computational overhead. Still, with bursty traffic patterns, this class of protocols suffers high contention penalty and latency [4]. The common solution is to design for low channel utility by employing complex multichannel gateways increasing available bandwidth (e.g., [5]). This is less attractive for constructing a csLWPAN as incentivising users to operate complex and expensive gateways could be challenging. This suggest that for csLWPANs there is an additional constraint that base station nodes should be low-cost devices similar to the IoT device nodes themselves.

To meet this challenge, we propose a new MAC protocol, DQ-N, for csLPWANs that supports thousands of nodes from a single low-cost gateway node. The protocol is inspired by distributed queueing and LPDQ [6], [7]. DQ-N provides near-optimal channel utility and latency characteristics. It supports thousands of devices within the assigned network simultaneously and is immune to bursty traffic. These optimizations enable the creation of efficient csLPWANs using commercially available low-cost LPWAN radios.

We envision users deploying public access csLWPANs nationwide, allowing multiple heterogeneous IoT applications to share a common and secure data network. This will simplify the IoT application development cycle, reduce costs, improve application reliability, and enhance the user experience.

The contributions of our paper are:

1) PlanIt, a realistic IoT network simulation and planning tool using both geographic and demographic information,
2) csLPWAN coverage analysis for Pennsylvania showing most cities can achieve 99% connectivity from a single crowdsourced (randomly located) base station,
3) and DQ-N, a near-optimal channel utility and latency MAC protocol for LPWAN networks that supports thousands of nodes with a single base station.

The structure of paper is as follows: in Section II we summarize prior work; in Section III we present PlanIt, a realistic LPWAN network planning tool with results covering all cities in Pennsylvania; in Section IV we present our improved DQ-N protocol for csLPWAN communication, including evaluation against other relevant protocols; and in Section V we present conclusions and future work.

II. Related Works

Ingenu (previously On-Ramp Wireless) designed a proprietary protocol called Random Phase Multiple Access (RPMA) for wide area networks. RPMA is a variation of CDMA [8]. After selecting a transmission slot, the transmitter performs a random delay. As long as two transmissions do not arrive simultaneously, RPMA allows correct decoding [9]. As with other contention-based MAC protocols, when the network load increases collisions will be more frequent, causing a reduction in the available network bandwidth. Ingenu is actively deploying private RPMA networks and claims to be the world's largest IoT network provider [10].

The LoRa Alliance specifies LoRaWAN, a protocol designed for low-cost IoT networks [11]. The physical layer of LoRaWAN uses Chirp Spread Spectrum modulation (CSS). CSS is very useful to recover data from weak signals, increasing the effective range of LoRaWAN transmissions. However, the MAC layer is very lightweight and essentially implements pure-ALOHA with Listen-Before-Talk, resulting in low channel utility under high traffic load due to packet collisions. Commercial LoRaWAN radios are widely available and implement the LoRaWAN MAC in software. Replacing the LoRaWAN MAC layer while keeping the CSS physical layer is an attractive option for developing csLPWAN equipment.

Random-FDMA (R-FDMA) has been developed with the goal to minimize the manufacturing cost for IoT devices [5]. By using ultra-narrow band transmitters and sophisticated wideband base stations, R-FDMA allows each IoT device to transmit using a random frequency. Although the lack of contention resolution introduces the possibility of interference within the same channel when two users are transmitting simultaneously, R-FDMA does not impose any constraints on how the node chooses an operating frequency. It is typically set in manufacturing allowing relaxed oscillator stability constraints. R-FDMA depends on each ultra-narrow-band channel being lightly utilized and forgoes typical medium access control mechanisms. This design simplifies device nodes but requires a more sophisticated base station to receive the random frequency signals. A typical R-FDMA base station would be implemented using a software defined radio (SDR) to sample the available spectrum and then perform all RF signal processing in software. This requires significant bandwidth between the SDR and host processor and enough processing power to process the received signal in real-time. Until a low-cost base station platform is developed, R-FDMA is less attractive for csLPWANs.

In the distributed queuing (DQ) literature, there are several recent attempts to adopt DQ into the IoT domain, most notably LPDQ [7], [12]. Both simulation and real world experiments demonstrate the superiority of DQ over ALOHA-based MAC protocols such as CSMA in terms of latency and throughput. LPDQ also has its own time synchronization and frequency hopping mechanisms. However, LPDQ is designed for high-frequency, higher-bandwidth wireless links and is not intended for low-rate IoT networks. As we will demonstrate in this paper, in a low data rate environment, the protocol overhead adversely affects the channel utility. In addition, LPDQ suffers larger latency under bursty traffic, as discussed in Section IV. Finally, LPDQ only supports upstream packets, which may be unacceptable in some IoT applications.

Low-power IoT coverage planning and network simulation are emerging topics [13]–[16]. B. Reynders *et.al* simulated both the physical and MAC layer for low-power networks in a square arrangement to evaluate the packet delivery ratio. The simulation explores the difference between wideband spread spectrum (LoRa-like) and an ultra narrowband (Sigfox-like) networks [13]. M. Centenaro *et.al* deployed a LoRa network in the real world and proved the feasibility of complementing IoT networks with long-range radio links [14]. SCALECycle was designed to solve intermittent and varying coverage problem using a mobile agent actively collecting data [15]. Y. Al Mtawa *et.al* proposed an algorithm to identify the holes in an IoT deployment.

Among radio propagation based simulations, a number of models are proposed to provide path loss estimation used for higher network layer simulation, such as ITM, Hata, and ITWOM [17], [18]. S. Kasampalis *et.al* demonstrated that ITWOM, though not perfect, gives more accurate results than previous models within a radius of 20 Km [17]. However, to our knowledge, there is no known simulation using demographic information to generate the test points; test points have been generated at random from uniform distributions.

Finally, one novel solution to solve IoT connectivity problem is to utilize people's smartphone as a public gateway device, as proposed by T. Zachariah *et.al*. The so-called

Universal Gateway can be implemented using Bluetooth Low Energy (BLE) on personal smartphones to remove the restriction that one IoT device has to connect to one specific smartphone to communicate with the Internet. Financial or other incentives can be used to increase user participation. We borrow this idea and explore the participation rate needed to achieve regional network coverage in csLPWANs.

III. PlanIt, Realistic LPWAN Planning

Estimating wireless coverage is a challenging problem. The most basic approach is to assume that the wireless devices are uniformly distributed in a square and the terrain is flat with only free-space path loss [13]. These are obviously not realistic assumptions. Since many IoT devices are designed to assist or improve human activities, it is reasonable to assume that their deployment shares similar characteristics as local demographic information. For instance, a city is more likely to have a dense deployment of IoT devices than a rural region. Therefore, a uniform distribution of test points does not represent the realistic deployment of human-centric IoT devices, thus reducing the credibility of these network simulation results.

Our approach in PlanIt is to select potential IoT device locations (Section III-A) within a region that reflects the local demographic characteristics. From the generated locations, we randomly select a subset to be crowd-sourced gateway devices. Then we use the Irregular Terrain with Obstructions Model (ITWOM) 3.0 to compute the path loss from each device to every gateway (Section III-B). If there is a gateway device within the link budget of the radio, a network connection is possible. The link loss information can then be used in a network simulator to produce realistic packet errors.

A. Selecting IoT Device Locations

Given longitude x and latitude y, the probability of an IoT device is $P_{IoT}(x, y)$, where P_{IoT} is the joint probability function defining the probability of all IoT devices. For a given region Ω, $\int_\Omega P_{IoT}(x, y)dxdy = 1$. PlanIt can consider several factors that can affect P_{IoT}:

- population density, $\rho(x, y)$,
- geographic-related information, such as $g(x, y)$ for topological effects,
- and demographic-related information, I, such as the influence of average age or income on IoT device use.

For simplicity, we define the adverse effect of topography as

$$P_{gt}(x, y) = -\boldsymbol{G} \cdot \nabla g = -\boldsymbol{G} \cdot \langle \frac{\partial g}{\partial x}, \frac{\partial g}{\partial y} \rangle,$$

where $\boldsymbol{G}$ is a vector representing strength of the effect.

Therefore, unnormalized P_{IoT}, P'_{IoT} can be written as

$$P'_{IoT}(x, y) = \rho(x, y) - \boldsymbol{G} \cdot \langle \frac{\partial g}{\partial x}, \frac{\partial g}{\partial y} \rangle + I\rho(x, y)$$

Hence the probability function P_{IoT} is

$$P_{IoT}(x, y) = \frac{P'_{IoT}(x, y)}{\int_\Omega P'_{IoT}}$$

Although in the real world these functions are continuous given a large region and population, for computational simplicity, we compute discrete values over a small interval.

To generate random IoT test points based on the probability distribution P_{IoT}, we will use the tiles when computing P_{IoT}. For tile (x, y), P_{IoT} is locally uniform, hence we can easily generate random points in the tile. For those boundary tiles, additional caution should be taken as we might generate points outside the region. If it is the case, we can simply discard these points.

By correctly choosing $\boldsymbol{G}$ and d, we can realistically approximate the probability function describing the IoT device locations. Below is a demonstration of how we generate IoT locations given population and geographic information. We choose $\rho(x, y) = 50((x - 0.5)^2 + (y - 0.5)^2)^{-1}$, $\Omega = (0, 1) \times (0, 1)$, $g(x, y) = ((x - 0.3)^{1.5} + (y - 0.3)^{1.5})^{-1}$. For illustration purpose we choose $\boldsymbol{G} =< 20, 20 >$. The first maps in Figure 1 show the probability density map of ρ, $\boldsymbol{G} \cdot \langle \frac{\partial g}{\partial x}, \frac{\partial g}{\partial y} \rangle$, and P_{IoT}. These functions are chosen to model a situation where population is centered at a city nearby a mountain. As we can see, the raw population is centered at the city yet influenced by the existence of the mountain, as indicated the dark corner in P_{IoT} density map. As a result, the IoT device location map reflects the population distribution in the P_{IoT} density map.

To select points in the United States, PlanIt uses demographic data from the U.S. Census 2010 Summary File 1 (SF1) [19] to estimate the population density (ρ). Census data is provided in several hierarchical levels from course to fine-grained: State, County, Subdivision, Place, Tract, Block group, and Block. For csLPWAN planning, we want to estimate network coverage over entire census-designated places. These census-designated places include cities, towns, boroughs, districts, municipalities, and townships. To further refine the test points within a place, we use data from one level down the hierarchy and break each place into the underlying census tracts. Each census tract has a distinct population density that is used to bias the sampling of points within the place.

As a real-world example, we consider the city of Philadelphia, Pennsylvania. Philadelphia is the most populated city among Pennsylvania's 57 cities with 1.526 million inhabitants as of the 2010 census. The city covers 365 square kilometers and is divided into 381 census tracts. Each census tract is fully contained within the city. We use the point selection algorithm considering only population density without any other geographic-related information (topological or demographic effects) to generate 1,000 test points in Figure 2a. Visually we can see the test points are not uniformly distributed. Some census tracts have multiple test points while others have no test points at all. The distribution of land area and the population of the census tracts are shown in Figure 2b. From this we can see there are 10 census tracts with fewer than 1,000 inhabitants and the majority of census tracts are smaller than 3 square kilometers. The census tracts without any sample points require further investigation. The large area near $(-75.20°, 40°)$ contains Fairmount Park, a large

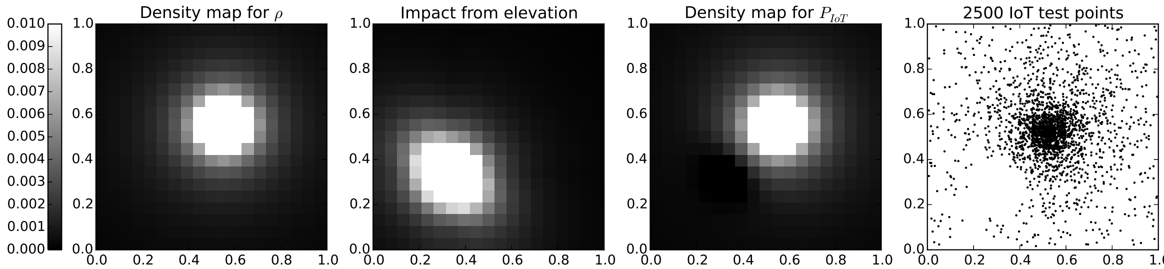
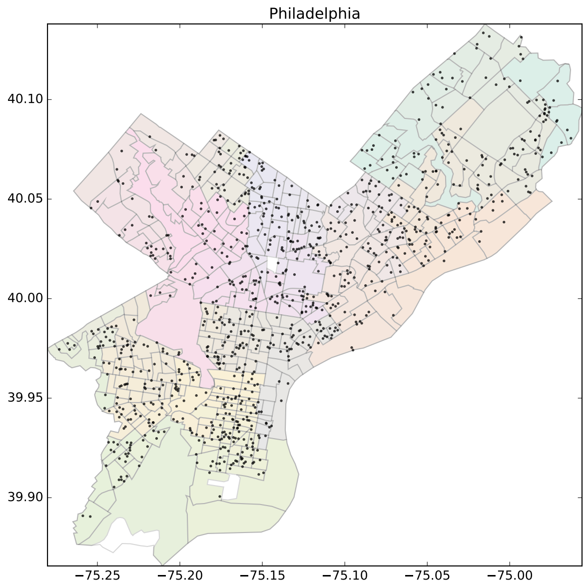
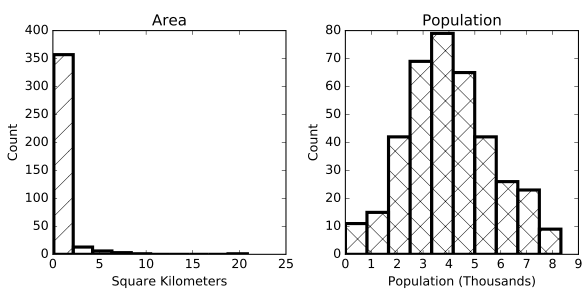


Fig. 1. Probability density map of $\rho(x, y)$, $\boldsymbol{G} \cdot \left\langle \frac{\partial g}{\partial x}, \frac{\partial g}{\partial y} \right\rangle$, and P_{IoT}. Notice the change of the map at bottom left corner due to geographic information. The right plot shows the 2500 IoT device locations generated based on P_{IoT}.



(a) Census tracts (colored) and test points (dots).



(b) Census tract area and population distributions.

Fig. 2. The data from 1,000 test points generated for the city of Philadelphia, Pennsylvania containing 381 census tracts.

urban park surrounding the Schuylkill River containing the Philadelphia Zoo and several other museums. The areas to the south are industrial zones with rail and shipyards to access the Delaware river. The area in the northeast contains the appropriately named Northeast Philadelphia Airport (the International airport, PHL, is outside of the city limits). With these facts in mind, the point distribution does closely follow the population distribution.

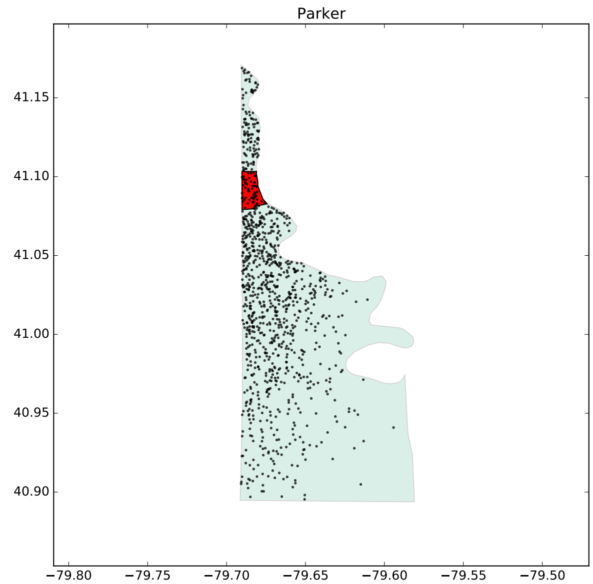


Fig. 3. The locations of 1,000 sample points generated for the city of Parker, Pennsylvania (red) and Armstrong County Census Tract 9503 (blue).

When considering other cities in Pennsylvania, we found that census tracts are not always fully contained within a city boundary as in Philadelphia. There are some cases when a city is contained within a census tract and the census tract is larger than the city. This is a common phenomenon when examining smaller census-designated places. The smallest city in Pennsylvania, Parker, falls into this category. In these cases, we select points within the census tract using a normal distribution with mean at the centroid of the city and standard deviation $\frac{1}{3}$ of the smallest distance from the city centroid to the bounding box created by the census tract, so there are at least three standard deviations to the nearest border. The goal is to generate points near the city center while still spanning the entire census tract with some probability. Any point generated outside of the census tract is rejected and a new random point is drawn. The resulting test points for Parker, Pennsylvania are shown in Figure 3. This shows the desired clustering of points near the city (red) while still representing the entire census tract (blue).

B. Path Loss Estimation

From a set of generated test points in a city or other region, we randomly select n points to be crowdsourced base stations. Because the test points were generated using population density and possibly other information, the selected base stations will also share this distribution. Then we compute the path loss from each point to all of the base stations using the Irregular Terrain with Obstructions Model (ITWOM) 3.0. This model improves on the Longley-Rice (ITM) model and estimates path loss taking into account topographic and ground clutter information. Although these models were developed for predicting DTV and FM broadcast coverage, they can also applied to the ISM frequency bands.

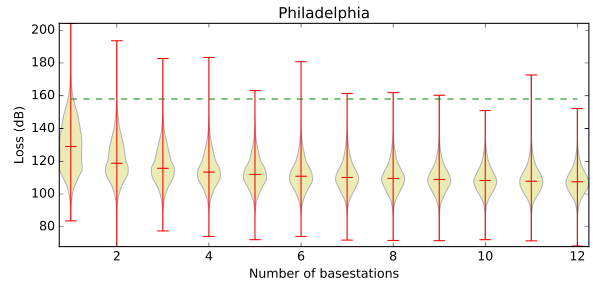
Topographic information for the model was obtained from the Shuttle Radar Topography Mission (SRTM) dataset with 1 arc-second resolution (approximately 30-meter resolution on the ground). This data has been previously shown to have good accuracy [20]. For each source/destination pair, the line of sight geodesic path is constructed by taking not more than 30-meter steps, to match the SRTM data resolution, from the source to the destination until reaching the destination. The final point along the path is always the destination. The elevation at each point is along the path is retrieved from the SRTM dataset. Before passing to the ITWOM library, if there are any negative elevation values, the entire path is shifted up by the absolute value of the most negative point so the minimum elevation is zero to prevent errors in the path loss calculation.

The complete set of ITWOM parameters are listed in Table I. The values for dielectric constant and conductivity are typical for a city environment. The transmitter height was set to 5 meters above ground level (AGL) which is easily reached from the roof of a 1-story building. The receiver height was set to 1 meter AGL to represent a device near ground level. Because of the reciprocity principle, it is irrelevant whether the IoT device or base station is the transmitter for path loss estimation.

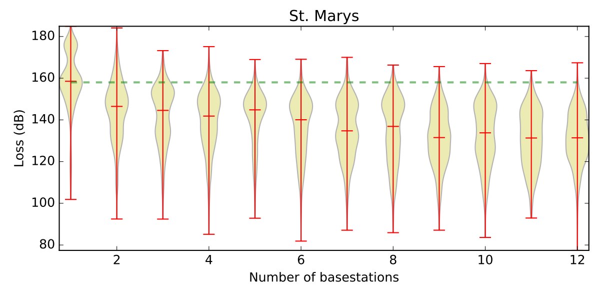
TABLE I
ITWOM PARAMETER VALUES USED TO COMPUTE PATH LOSS.

Parameter	Value
Transmitted Height	5 meters
Receiver Height	1 meter
Earth Dielectric Constant	5.0 (city)
Earth Conductivity	0.001 (city)
Atmospheric Bending Constant	301.0
Frequency	900 MHz
Polarization	Horizontal
Location Variability	50%
Time Variability	50%

Using these parameters, we generate 100 sets of 1,000 test points and randomly select $\{1 \dots b\}$ points from each set to act as base stations, then compute the path loss from all points to every base station keeping only the lowest loss



(a) Path loss in Philadelphia, Pennsylvania.



(b) Path loss in St. Marys, Pennsylvania.

Fig. 4. The path loss distribution from 100 rounds of evaluating 1,000 test points with randomly assigned base station nodes in two Pennsylvania cities. The dashed line at 158 dB indicates the threshold for reception.

path. Figure 4a shows the path loss distribution for Philadelphia, Pennsylvania along with a hypothetical 158 dB receive threshold. This threshold was selected as a fair estimate from available LPWAN radio specifications which vary from 149 to 175 dB. This shows that a single randomly located base station has a very good chance of providing connectivity to most of Philadelphia with median path loss of 130 dB. This result is somewhat surprising and we should keep in mind that ITWOM only considers the elevation data measured from space provided by the SRTM dataset. It is possible that the reported elevations reflect the top of buildings, especially in a dense city, resulting in urban canyon effects on the surface. However, this result gives us confidence that csLWPANs can cover a large population using very few base stations.

Analyzing other cities in Pennsylvania yields similar results to Philadelphia. Figure 5 shows the estimated connectivity using one base station and a 158 dB receive threshold for every city in Pennsylvania. The median connectivity is 99.2% and in all but two cities the median connectivity is greater than 90%. The worst performing city in Pennsylvania is St. Marys, shown in Figure 4b. Investigating St. Marys reveals although classified as a city, it is a geographically large area of nearly 100 square miles in the Allegheny Mountain region spanning an elevation range of more than 250 meters. For comparison, Philadelphia covers 141 square miles but only has an elevation range of 100 meters with more gradual elevation changes. These two factors combine to yield a challenging region to cover.

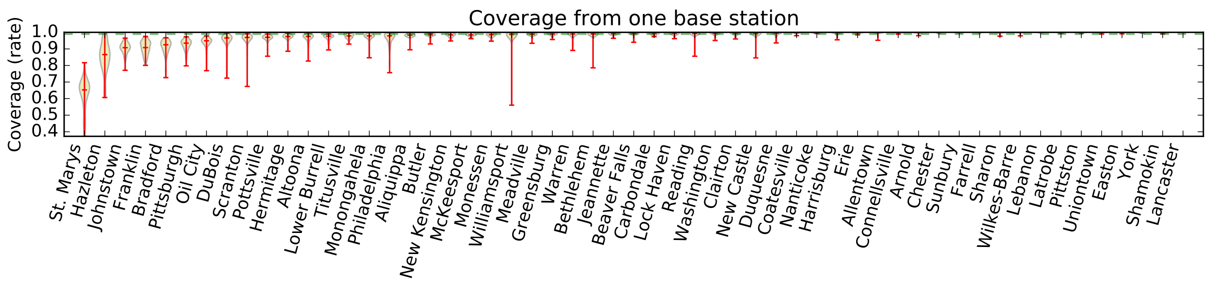


Fig. 5. The coverage rate for all cities in Pennsylvania with one base station and a receive threshold of 158 dB. The median connectivity, 99.2%, is indicated by a dashed line.

IV. DQ-N FOR CSLPWANS

From our work with PlanIt (Section III), we believe csLWPANs are a promising approach to provide low-rate IoT network coverage utilizing very few base stations. In the extreme case, 98% of Philadelphia's 1.526 million inhabitants could connect to a single base station. For a low data rate network with simple base stations, basic connection overhead alone could cripple the network with this many users. Our approach is to reconsider the use of contention-based media access control protocols in favor of a distributed queuing-based approach that we call DQ-N that to better support highly utilized networks.

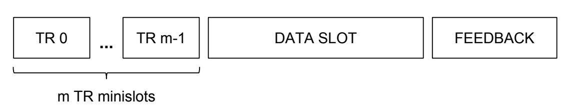
Distributed queuing (DQ) is a hybrid media access control protocol where the coordinator broadcasts contention-free transmission queue values to individual devices in response to contention-based transmission requests. In contrast to other media access approaches, the coordinator maintains minimal information about the network and is only responsible for processing and broadcasting the results of the contention resolution process. Each device in the network maintains two queue lengths, namely the contention resolution queue (CRQ) and data transmission queue (DTQ). Using only this information, devices can compute contention-free transmit times in a fully distributed fashion. Although DQ can be implemented in the frequency domain, this paper focuses on the time domain as we plan to exploit different frequency channels to increase network capacity.

To request a data slot using DQ, the device independently chooses a minislot from $\{0, 1, \ldots, m-1\}$ at random and then sends a TR at that chosen minislot in the next frame. Since there may be many devices attempting to transmit simultaneously and there are a fixed number of minislots, a contention may occur at that chosen minislot. The base station provides contention information in the feedback. If contention occurs in a TR minislot, each contending device enters the CRQ. If a TR is successful (no contention) the device enters the DTQ. To calculate the device's CRQ position, the minislot states provided in the feedback are scanned from lowest index to highest index. If a contention is indicated in a occurs in a minislot, the device increases the current CRQ length by 1. The scanning process continues until the device reaches its requested minislot index. The value obtained by this process is the position of the device in the CRQ. Devices in the CRQ backoff for the computed CRQ number of frames and repeat the TR process. Analogously, the position of DTQ is calculated by scanning success states in the feedback and incrementing the base DTQ value accordingly. Devices in the DTQ wait for the indicated data slot and then communicate without contention.

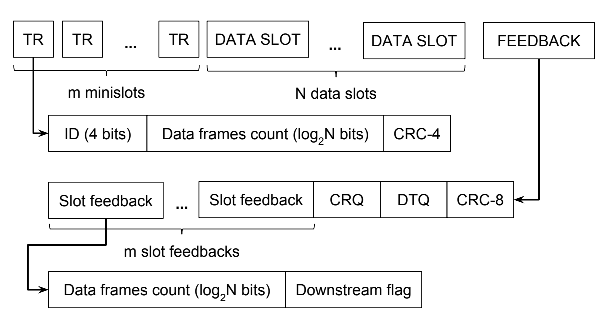
An important feature of DQ is that once the CRQ and DTQ values are computed, the device can switch to sleep mode to save energy and wake up at the scheduled time to transmit data. There is no need to continuously sense the network traffic load. This is an advantage to the IoT domain where channel sensing typically consumes significant device energy [21]. If a device detects that both CRQ and DTQ are empty, it may use any unused data slot, accepting the possibility of contention. This behavior is similar to slotted ALOHA and reduces the latency significantly when the network load is low.

Figure 6a shows the DQ frame structure where the channel is divided into multiple fixed-length frames containing: 1) m minislots for transmission requests (TRs), 2) a contention-free data transmission slot, 3) and a feedback slot.

Because current LPWAN radios use relatively small packet



(a) DQ frame structure.



(b) DQ-N frame structure.

Fig. 6. The frame structure for DQ and DQ-N.

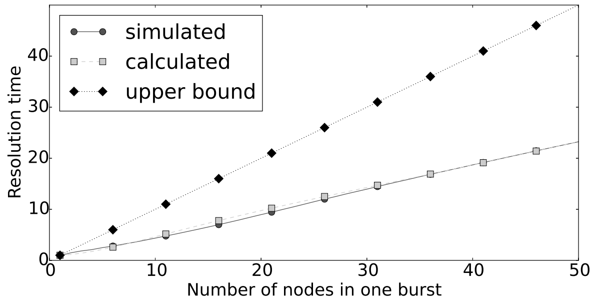


Fig. 7. DQ transmission request burst resolution time, L_n, with $m = 3$.

sizes, the data slot size is constrained by the maximum radio packet size in practice. As the TR and feedback messages are overhead, the overall protocol efficiency with small data slots is low. The goal of DQ-N is to support N data transmission slots between each TR and feedback message. Figure 6b shows the modified DQ-N frame structure. As a result of this change, the overall protocol efficiency can be tuned by varying N.

The benefit of DQ-N is that contention is reduced by the coordinator broadcasting feedback containing the current CRQ and DTQ length as well as TR results. The TR results consist of an array of states for each of the minislots, namely, *idle*, *success*, and *contention* and the number of data slots requested. This information is used to compute the queue position at each node by scanning the results for each TR in the same way as in DQ but now incrementing the DTQ queue length by the number of requested data slots. If a device wants to minimize latency in a network with multiple base stations, before transmitting a TR, it can first sense the load of each base station by receiving a feedback message containing the CRQ and DTQ lengths. The device can then send it's TR to the channel with shortest queue length.

To enable downstream messages in DQ-N, we add a flag to the feedback message to indicate if there are downstream packets pending for a device after a successful TR. After seeing this flag, the device will complete the upstream transmission and then request a downstream data slot by transmitting a receive request (RR) message in a random minislot. The RR protocol is the same as the TR, however, at the assigned data slot the base station transmits the data. Additionally, at any time a device can send a receive request (RR) message to contend for downstream data slots even if it has not seen a downstream flag.

A. Analysis

The contention resolution algorithm of DQ is a tree splitting algorithm. All devices that transmit a TR in the same minislot will compute the same CRQ value and hence occupy a common branch in the contention tree. We can use this to calculate the expected waiting time in the system for a transmission burst. Figure 7 shows the expected number of frames required to resolve a burst of transmission request versus the number of nodes in the burst using DQ with $m = 3$. This is an important property for IoT applications as it improves the stability of the system during bursty loads.

For further analysis, since DQ-N supports requesting multiple data frames in each TR, it is reasonable to assume the total number of requested data slots per TR forms a distribution with mean value λ and standard deviation σ_a. Accordingly, we compare DQ-N to an M/G/1 queue with input ratio $\rho = \lambda/N$. The following list defines constants used in the analysis of DQ-N:

- ρ : server utilization (traffic load).
- N : number of data slots per frame.
- m : the number of minislots per frame.
- λ : average number of data slots per TR.
- σ_a : standard deviation of data slots for each TR.
- σ_s : standard deviation of service time.
- γ : average number of TR per frame.

By Little's Law and the Pollazcek-Khintichine formula, the average delay time in DQ-N based $M/G/1$ queue is:$W_{M/G/1} = \frac{L_q}{\lambda} + \frac{\lambda}{N} = \frac{\lambda^2\sigma_s^2+\rho^2}{2\lambda(1-\rho)} + \frac{\lambda}{N}$, where $\sigma_s^2 = \sigma_a^2/N^2$. If we know the the expected value of λ and σ_a is relatively small compared to N, we can choose the number of mini-slots based on λ, as given in Theorem 1.

Theorem 1. *It takes no more than n frames to resolve λn TRs if $m \geq \gamma + 1$.*

Proof. Since each unit time the server can process N data slots, one can show that it is sufficient to prove that $L_{(\lambda n+\lambda)/N} \leq 1 + L_{(\lambda n)/N}$. Hence it is equivalent to show $L_{n+1} \leq \frac{N}{\lambda} + L_n$. According to DQ theory we need $\frac{m^{n-1}}{m^n-1} \leq \frac{\lambda}{N} = \frac{1}{\gamma}$. Since $\frac{m^{n-1}}{m^n-1}$ is monotonically decreasing, and $m \geq \gamma + 1$, $\frac{m^{n-1}}{m^n-1} \leq \frac{1}{\gamma}$, as desired. □

B. Simulation

To simulate the performance of DQ-N, we first use PlanIt to generate device locations in three counties in central Pennsylvania, namely Union, Northumberland, and Snyder County. The test area covers 2,976 km^2 with a population of 180 thousand. 5,000 locations are randomly drawn based on the procedure discussed earlier. We filter out any locations with a negative signal to noise ratio (SNR), as those devices are unlikely to transmit to or receive from any valid information with the base station. We also assume each device communicates to the base station with the lowest path loss. The path loss values are converted to bit error rate (BER) for use in simulation using the following equations.

$$SNR = TI + GT - NJ - L + G - NF$$
$$\frac{E_b}{N_0} = SNR - 10log_{10}(\frac{f_b}{B})$$
$$BER = \frac{1}{2}e^{-\frac{1}{2}\frac{E_b}{N_0}}$$

The meanings and values of the parameters are listed in Table II.

For the base station transmitter, we use $\frac{1}{2}$ Watt as the transmitter power and 200 ms for the duration of the time slot, to comply with FCC regulations for frequency hopping systems in the 900 MHz band. We select 1200 bps datarate and 12.5 KHz channel bandwidth to improve receiver sensitivity and therefore increase the effective range. The remaining transceiver parameter values are obtained from the TI CC1120 and CC1190 narrowband transceiver and range extender datasheets [22].

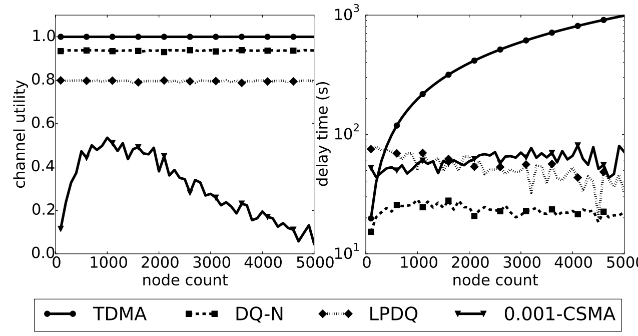
TABLE II
PARAMETERS USED TO CALCULATE BER

Name	Value
RF power delivered to the transmitter (TI)	-3 dBW
Transmitter antenna gain (GT)	3 db
Johnson Noise (NJ)	-114 dbW
Path loss (L)	Computed by PlanIt
Receive antenna gain (G)	3 db
Receiver noise figure (NF)	7 db
Channel data rate (f_b)	1200 bps
Channel bandwidth	12.5 KHz
PHY model	BPSK

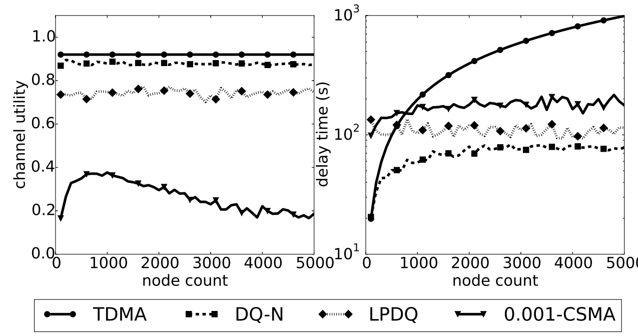
For comparison, we also simulate P-persistent CSMA, ideal TDMA, and LPDQ in addition to DQ-N. These protocols are chosen because they form the basis for many popular MAC layer protocols. P-persistent CSMA operates like traditional CSMA but a node only transmits on idle channels with probability P (we use $P = 0.001$) to reduce contention. Ideal TDMA assumes a static round-robin transmit schedule for all nodes. LPDQ is a previous implementation of distributed queuing for low-power wireless networks. DQ-N improves upon LPDQ by scheduling multiple data slots per frame resulting in reduced protocol overhead for the short frame sizes common in LPWAN systems and supporting both upstream and downstream traffic.

We performed two different simulations, the ideal case without packet loss and a realistic case with packet loss caused by the path loss predicted by PlanIt. For both cases, we generate sufficient upstream traffic to saturate the network evenly distributed among all of the nodes (e.g., if there are 100 nodes and 1,200 bps available upstream bandwidth, each node would generate 12 bps of upstream traffic) and then sample the network utility and latency after a warm up period. Network utility is defined as the ratio of successfully received data without contention or corruption versus the available bandwidth. The latency is defined as the time from when a node begins to send a packet to the time when a base station successfully receives it, thus including the contention and waiting time. No re-transmissions are used in the simulations as such effort will confound the simulation results and typically belongs to upper layer protocols. We also measure the duty cycle for each node to demonstrate the energy intensity of each protocol.

Transmitted packet sizes are selected from a normal distribution with a mean of 240 bytes and standard deviation of 120 bytes. Since the total available bandwidth is 1,200 bps and the maximum time slot is 200 ms, each data slot contains 30 bytes of data. We choose $N = 16$ and $m = 8$ for the DQ-N parameters. We also subtract the protocol overhead when calculating maximum channel utility, therefore use 0.94 for DQ-N and 0.8 for LPDQ which saturates the network using these parameters.



(a) Simulation results for the ideal environment.



(b) Simulation results for the realistic environment.

Fig. 8. LPWAN network simulations for saturated networks.

Figure 8a shows the simulation results for an ideal environment where there is no packet loss. In this case, DQ-N achieves similar performance to LPDQ but demonstrates lower latency. However, since DQ-N reduces protocol overhead, it has higher channel utility.

Figure 8b shows the simulated utility and delay considering packet loss. The latency for all protocols except TDMA increases when compared to the idea case. The channel utility for all protocols is reduced due to packet loss. For DQ-N and LPDQ, in addition to lost data messages, a lost TR or feedback message may result in an unscheduled data slot, further reducing utility. However, DQ-N still outperforms LPDQ and P-CSMA in terms of channel utility and latency.

Figure 9 shows the radio duty cycle distribution in the realistic environment with 2,500 nodes. Although TDMA exhibits the optimal duty cycle $(1/2500)$, it is not practical as the number of nodes has to be fixed and dynamically changing the TDMA schedule requires a more complex protocol. Nevertheless, the simulation shows the improvement of DQ-N over LPDQ as it eliminates the need for nodes to make multiple TRs for messages longer than one data slot.

V. CONCLUSION AND FUTURE WORK

In this paper, we consider crowdsourced low-power wide-area networks (csLPWANs). A csLPWAN is a LPWAN where

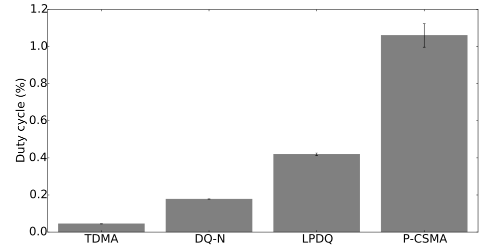


Fig. 9. Radio duty cycle in the realistic environment.

a subset of users provide base stations without coordinating the location of the base stations. To better understand csLPWAN behavior, we have created a csLPWAN network planning tool, PlanIt. PlanIt combines topographic and demographic information to give a more realistic representation of real-world csLPWAN connectivity. Using PlanIt, we found the median connectivity over all cities in Pennsylvania was 99.2% from a single randomly located base station within each city. Using the path loss data from PlanIt, we have also simulated and compared the efficiency of different LPWAN protocols. To simplify csLPWAN gateways and improve network utilization, we have designed and analyzed DQ-N, an extension of the DQ protocol optimized for highly utilized low-rate csLPWAN networks. By presenting mathematical analysis and numeric simulation results, we show the superior performance of DQ-N and demonstrate the ability of a single DQ-N base station to support thousands of nodes in a csLPWAN. We believe these results will help catalyze the deployment of future csLWPAN networks. We are currently developing a low-cost base station to support networks of thousands of LoRa devices using the DQ-N protocol. More information on this work and a web-based version of PlanIt are available at `http://cslpwan.me`.

References

[1] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (IoT): A vision, architectural elements, and future directions," *Future Generation Computer Systems*, 2013.

[2] N. Oyj, "LTE evolution for IoT connectivity," *Nokia Corporation White Paper*, 2016.

[3] A. Laya, C. Kalalas, F. Vazquez-Gallego, L. Alonso, and J. Alonso-Zarate, "Goodbye, ALOHA!" *IEEE Access*, 2016.

[4] B. Nardelli and E. W. Knightly, "Robust CSMA: Adapting to channel and traffic asymmetry," in *Proceedings of Sensing, Communication, and Networking*, 2015.

[5] M. T. Do, C. Goursaud, and J. M. Gorce, "On the benefits of random FDMA schemes in ultra narrow band networks," in *International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, 2014.

[6] W. Xu and G. Campbell, "A distributed queueing random access protocol for a broadcast channel," in *Proceedings of Communications Architectures, Protocols and Applications*, 1993.

[7] P. Tuset-Peiro, F. Vazquez-Gallego, J. Alonso-Zarate, L. Alonso, and X. Vilajosana, "Experimental energy consumption of frame slotted ALOHA and distributed queuing for data collection scenarios," *Sensors*, 2014.

[8] T. J. Myers, D. T. Werner, K. C. Sinsuan, J. R. Wilson, S. L. Reuland, P. M. Singler, and M. J. Huovila, "Light monitoring system using a random phase multiple access system," 2013.

[9] G. Margelis, R. Piechocki, D. Kaleshi, and P. Thomas, "Low throughput networks for the IoT: Lessons learned from industrial implementations," in *IEEE 2nd World Forum on Internet of Things*, 2015.

[10] J. Schneider and K. Garvin, "Ingenu announces launch of the machine network," *Ingenu Press Release*, 2015.

[11] N. S. (Semtech), M. L. (Semtech), T. E. (IBM), T. K. (IBM), and O. H. (Actility), "LoRaWAN specification," LoRa Alliance, Tech. Rep., January 2015.

[12] P. Tuset-Peiro, F. Vazquez-Gallego, J. Alonso-Zarate, L. Alonso, and X. Vilajosana, "LPDQ: A self-scheduled TDMA MAC protocol for one-hop dynamic low-power wireless networks," *Pervasive and Mobile Computing*, 2015.

[13] B. Reynders, W. Meert, and S. Pollin, "Range and coexistence analysis of long range unlicensed communication," in *23rd International Conference on Telecommunications*, May 2016.

[14] M. Centenaro, L. Vangelista, A. Zanella, and M. Zorzi, "Long-range communications in unlicensed bands: the rising stars in the iot and smart city scenarios," *CoRR*, 2015.

[15] Q. Zhu, M. Y. S. Uddin, Z. Qin, and N. Venkatasubramanian, "Upload planning for mobile data collection in smart community internet-of-things deployments," in *2016 IEEE International Conference on Smart Computing*, May 2016.

[16] Y. A. Mtawa, H. S. Hassanein, and N. Nasser, "Identifying Bounds on Sensing Coverage Holes in IoT Deployments," in *IEEE Global Communications Conference*, Dec 2015.

[17] S. Kasampalis, P. I. Lazaridis, Z. D. Zaharis, A. Bizopoulos, S. Zetlas, and J. Cosmas, "Comparison of Longley-Rice, ITM and ITWOM propagation models for DTV and FM broadcasting," in *International Symposium on Wireless Personal Multimedia Communications*, June 2013.

[18] M. Hata, "Empirical formula for propagation loss in land mobile radio services," *IEEE transactions on Vehicular Technology*, 1980.

[19] US Census Bureau, "United states census," 2010.

[20] P. Guth, "Geomorphometric comparison of ASTER GDEM and SRTM," in *A special joint symposium of ISPRS Technical Commission IV & AutoCarto in conjunction with ASPRS/CaGIS*, 2010.

[21] M. A. Razzaque and S. Dobson, "Energy-efficient sensing in wireless sensor networks using compressed sensing," *Sensors*, 2014.

[22] Texas Instruments, "CC1120 High-Performance RF Transceiver for Narrowband Systems," *Datasheet Available Online at: http://www. ti. com/lit/ds/symlink/cc1120. pdf*, 2015.