Reporting Mechanisms for Internet of Things

Chia-Wei Chang, Yi-Bing Lin, and Jyh-Cheng Chen(✉)

Department of Computer Science, National Chiao Tung University, Hsinchu 300, Taiwan
{chwchang,liny,jcc}@cs.nctu.edu.tw

Abstract. Energy saving is one of the most important issues for Internet of Things (IoT). An intuitive way to save energy of IoT devices is to reduce the reporting frequency to the IoT server. However, to do so, the time-variant values are distorted, which may be influential to the measured results. In this letter, we take PM2.5 application as an example to discuss the relation between energy efficiency and data accuracy. Through analyzing PM2.5 concentration collected via LoRa at National Chiao Tung University (NCTU) from 2016 to present, two reporting mechanisms based on timer and threshold, respectively, are proposed. The experimental results demonstrate that the threshold-based reporting outperforms the timer-based reporting by more than 37% in energy saving when the accuracies of these two reporting mechanisms are the same.

Keywords: Internet of things · LoRa · Reporting frequency · Data accuracy · Packet loss · PM2.5

1 Introduction

In the recent years, many cities utilize the environmental information obtained from small sensors to improve living experience. Small sensors collect data and forward them to the server via wireless technologies such as NB-IoT [1], LoRa [2], and SigFox [3]. After that, the server can analyze the data and make strategic decisions based on the analytical results. For instance, an intelligent ventilation system can dynamically control intake/extractor fans based on the amount of carbon dioxide emission sensed in a classroom. By improving the air quality, students can breathe fresh air which makes students doze off less and concentrate more in the class.

To develop a smarter campus, NCTU is deploying several IoT-based applications and devices to monitor temperature, particulate matter (PM), carbon dioxide emission, etc. For those devices deployed in hard-to-reach locations, replacing batteries for these devices is a big issue. One could prolong the life time of a device by letting the device reduce the number of communication (i.e., reduce the reporting frequency to the IoT server). The device then could enter the sleeping mode to save their energy [4,5].

J.-L. Chen et al. (Eds.): WiCON 2018, LNICST 264, pp. 193–202, 2019.
https://doi.org/10.1007/978-3-030-06158-6_20

Nevertheless, environmental data is time variant, where every trail of environment changes may be necessary for IoT server to conclude a significant discovery. It is unwise to trade all trails of changes for saving more energy. The questions are how many trails and which trails to be traded. Therefore, in this letter, two reporting mechanisms that extend the battery life of the IoT devices are investigated. The tradeoff between energy efficiency and accuracy of the collected data are discussed. To our best knowledge, most PM2.5 studies focus on forecasting the degree of air pollution, dealing with large scale data, or searching for the sources of air pollution which do not cover the issue stated above [6–9]. This letter compares the reporting mechanisms for the data accuracy and power saving. The main contributions of the letter are as follows: (a) we establish an IoT-based application on our campus including sensor device deployment, wireless system establishment, and IoT-based platform construction (front-end and back-end systems), (b) we continuously collect massive PM2.5 data which is gathered from Aug 2016 to present, (c) we observe the tradeoff between data accuracy and reporting frequency with various interesting finding, and (d) we improve the data accuracy and energy efficiency.

The letter is organized as follows: Sect. 2 presents two reporting mechanisms, and some investigations on data accuracy are given. Section 3 compares the performance of these two reporting mechanisms. Experimental results are provided. Finally, Sect. 4 concludes several interesting findings.

(a) PM2.5 sensor device. (b) LoRa gateway. (c) Locations of devices and gateways.

Fig. 1. The PM2.5 measurement system on the NCTU campus.

2 Reporting Mechanisms

In this section, two reporting mechanisms to collect the sensed data from the IoT devices are investigated: (1) *timer-based reporting*, and (2) *threshold-based reporting*. Here, we use PM2.5[1] concentration measurement deployed on the NCTU campus with 11 rooftop PM2.5 sensors as an example to evaluate our proposed reporting mechanisms [10]. The collection of PM2.5 concentration reported per minute since August 2016 is also used for cross validation [10]. Figure 1 shows eleven IoT devices and two IoT servers deployed on the campus. Every IoT

[1] PM2.5 is the suspended particulate matter smaller than 2.5 µm in diameter.

device is a PM2.5 sensor (PMS700) and a LoRa module (GL6509)[2]. The sensor is used to measures PM2.5 concentration and the LoRa module is used to transmits measurement results to the IoT server (see Fig. 1(a)). The IoT server is a LoRa outdoor 915 MHz gateway (WAPS-232N, see Fig. 1(b)) with two antennas where the receiver sensitivity can be down to −142 dBm and the RF TX power is 0.5 W (up to 27 dBm) [11].

Table 1. List of parameters

T	The obvervation time period
τ	The timer interval for timer-based reporting
ϵ	The threshold for threshold-based reporting
v_i	The PM2.5 value measured at time t_i
v_I	The latest reported PM2.5 value measured by the IoT device before t_i
V	The set of reported PM2.5 values measured by the IoT device during the observation period T
V_τ	The set of reported PM2.5 values measured per τ period
V_ϵ	The set of reported PM2.5 values where $\|v_i - v_j\| > \epsilon$ for $v_j \in V_\epsilon$
$\|\|V\|\|$	The number of reported PM2.5 values in V
ε_τ	The expected error with timer interval τ
ε_ϵ	The expected error with threshold ϵ
f_τ	The reporting frequency with timer interval τ
f_ϵ	The reporting frequency with threshold ϵ

In our study, the *reporting frequency* (f) and the *expected error* (ε) are used as the metrics to evaluate the energy efficiency and the accuracy of collected PM2.5 concentration, respectively. Their definitions are given as follows and are summarized in Table 1. An IoT device (a PM2.5 sensor) measures the PM2.5 values and reports them to the IoT server. Let V be the set of PM2.5 values measured by the IoT device during an observation time period T, which are reported to the IoT server. Then the reporting frequency f is defined as:

$$f = \frac{||V||}{T} \tag{1}$$

Let v_i be the PM2.5 value at time t_i. Let $V(t_i) = \{v_j \in V | t_j \leq t_i\}$ and $t_I = \max_{v_j \in V(t_i)} t_j$. Then v_I is the PM2.5 value reported to the IoT server at the latest moment before t_i. In the measured system, we consider the PM2.5 value at t_i as v_I. Therefore, the reporting accuracy at t_i is determined by the measured error

[2] The maximum working power for PMS7003 is 100 mA × 5 V= 0.5 W at most. The RF power with 20dBm for GL6509 is 0.45 W. In other words, almost 50% of the device's energy is consumed by the LoRa transmission.

$\varepsilon(t_i)$ defined as $\varepsilon(t_i) = \frac{|v_i - v_I|}{|v_i|}$, and the reporting accuracy for V is determined by the expected error ε expressed as:

$$\varepsilon = \int_{t_i=0}^{T} \frac{\varepsilon(t_i)}{T} dt_i \tag{2}$$

In the timer-based reporting, the IoT device periodically sends the measured PM2.5 values with the fixed intervals τ. Let V_τ be the set of the PM2.5 values reported to the IoT server during T. Then from Eq. (1), the reporting frequency f_τ of the timer-based reporting mechanism with the interval τ is:

$$f_\tau = \frac{||V_\tau||}{T} = \frac{1}{\tau}$$

and the measured error ε_τ is the ε value computed for V_τ using Eq. (2).

In the threshold-based reporting with the threshold ϵ, the IoT device sends the PM2.5 value at t_i to the IoT server if $|v_i - v_I|$ is larger than the threshold ϵ (i.e., $|v_i - v_I| > \epsilon$). The reporting frequency f_ϵ of the threshold-based reporting mechanism with the threshold ϵ is:

$$f_\epsilon = \frac{||V_\epsilon||}{T}$$

and the measured error ε_ϵ is the ε value computed for V_ϵ using Eq. (2).

Figure 2 depicts the measured PM2.5 values at location 'A' from December 2016 to January 2017 where PM2.5 concentration was sampled every minute (62 days $\times$ 24 h/day $\times$ 60 min/hr = 89,280 samples). That is, the unit for t_i is *minute*. These sampled values v_i will be utilized for both timer-based and threshold-based reporting mechanisms in the following sections[3].

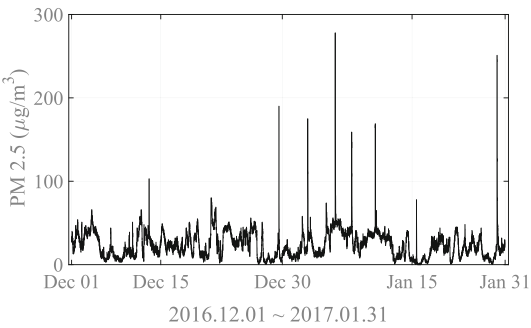
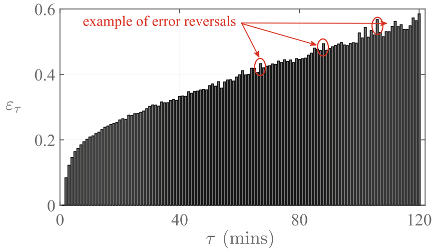


Fig. 2. Histogram of PM2.5 data at location 'A'.

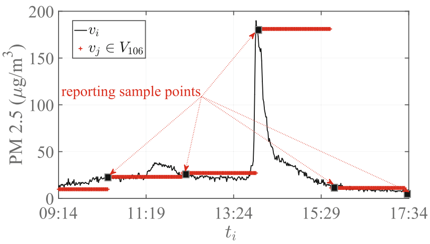
[3] The observation period lasted 4 months with 174,240 sampled data (121 $days \times 24\ h \times 60\ min$). Figure 2 illustrates partial results.

2.1 Timer-Based Reporting Mechanism

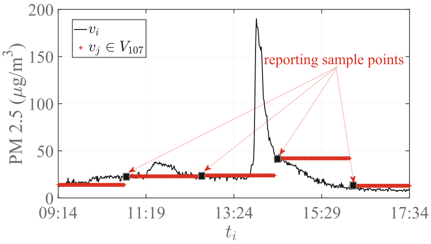
The timer-based reporting mechanism reports the measured data every moment when the timer τ is expired (Zero-Order Hold [12]). Figure 3(a) demonstrates that most expected errors ε_τ gradually ascend with the increase of τ. A smaller τ expires sooner than a larger τ and causes a larger $||V||$, which in general makes the data reporting to the server more accurate but consumes more energy. What is interesting is that several **error reversals**[4] appear, e.g., when $\tau = 64$ $(mins)$, $\tau = 88$ $(mins)$, and $\tau = 106$ $(mins)$ as shown in Fig. 3(a).



(a) Expected error ε_τ vs timer interval τ.



(b) τ = 106 *mins*.

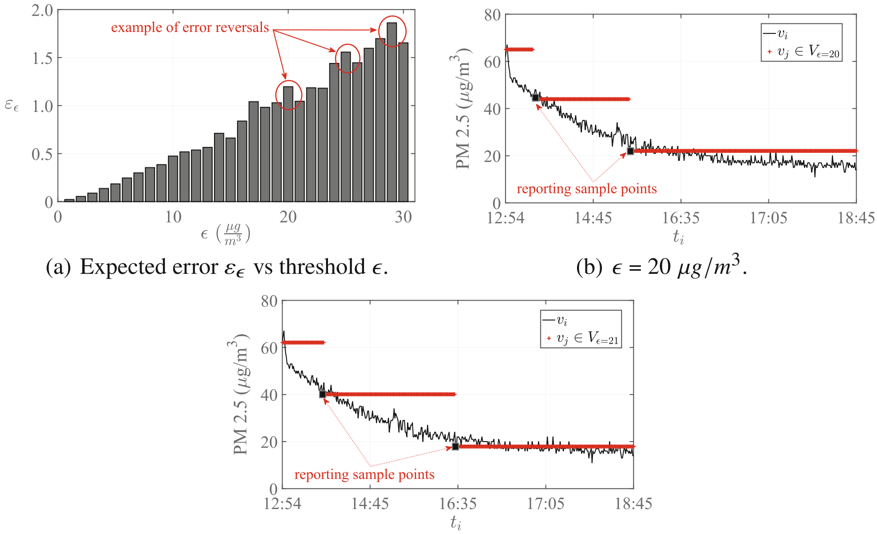


(c) τ = 107 *mins*.

Fig. 3. Timer-based reporting. (b) and (c) are extracted from Fig. 2 during 9:14 $\sim$ 17:34 on Dec. 29, 2016.

The error reversal phenomenon can be explained in Figs. 3(b), (c) by re-examining the τ configured as 106 $(mins)$ and 107 $(mins)$ extracted from *9:14 a.m.* to *5:34 p.m.* in the day of Dec. 29 of Fig. 2. When τ is set to 106 $(mins)$ in Fig. 3(b), the reporting moment around 190 $(\frac{\mu g}{m^3})$ resulting large ε_{106} in the following 106 min. On the other hand, the sample points shown in Fig. 3(c) do not lead to large differences between v_i and v_I when τ is set to 107 $(mins)$. Even though the process that the PM2.5 concentration reaches the highest value and then slowly goes down for almost 2 h (i.e., *13:24∼15:29*), it is a relatively high variation as far as large τ is concerned.

[4] *Error reversal* is defined as an incident where the reporting accuracy obtained by a smaller timer is worse than that by a larger timer.



(a) Expected error ε_ϵ vs threshold ϵ. (b) $\epsilon = 20\ \mu g/m^3$. (c) $\epsilon = 21\ \mu g/m^3$.

Fig. 4. Threshold-based reporting. (b) and (c) are extracted from Fig. 2 during 12:54 ∼ 18:45 on Jan. 25, 2017.

Observation 1: The expected error ε does *NOT* always increase by enlarging the timer interval τ. In some cases, reducing τ counterintuitively increases the expected error ε due to the error reversal phenomenon.
Observation 2: For v_j in V, when f is low, changes of v_i are relatively dramatic. In this case, the number of error reversals is greater in a time period with more severe fluctuation of PM2.5 values.

2.2 Threshold-Based Reporting Mechanism

In the threshold-based reporting, the PM2.5 sensor sends the data only on critical changes of data values. That is, when the difference between the latest reported data and the current measured data exceeds a certain threshold ϵ. Figure 4(a) depicts the relation between the threshold ϵ and the expected error ε where ε is enlarged as the ϵ is augmented. Similarly, some threshold sizes, e.g., when $\epsilon = 20$ ($\frac{\mu g}{m^3}$), $\epsilon = 25$ ($\frac{\mu g}{m^3}$), $\epsilon = 29$ ($\frac{\mu g}{m^3}$), have larger ε than their successors, i.e., $\epsilon = 21$ ($\frac{\mu g}{m^3}$), $\epsilon = 26$ ($\frac{\mu g}{m^3}$), and $\epsilon = 30$ ($\frac{\mu g}{m^3}$), respectively. The reason causing error reversals in the threshold-based reporting mechanism is the same as that in the timer-based reporting mechanism (see Figs. 4(b), 4(c)).

Observation 3: The expected error ε_ϵ does *NOT* always increase by enlarging the threshold ϵ.

Observation 4: Error reversal phenomenon typically occurs when PM2.5 values instantly ascend/descend and then keep in similar values for a long time. For the reporting point of a relatively smaller τ or ϵ is located at the beginning of ascending/descending PM2.5 values. On the other hand, the reporting point of a larger τ or ϵ has a value closer to the upper/lower plateau.

3 Performance Discussion

In this section, we investigate the accuracy performance of the timer-based and threshold-based reporting mechanisms with and without considering packet loss.

3.1 Performance Without Packet Loss

Figure 5 demonstrates that when the expected errors of both reporting mechanisms are fixed at 0.5, the threshold-based reporting ('+' marker) outperforms the timer-based reporting ('□' marker) approximately by more than 37% in energy saving. This phenomenon is due to the fact that the threshold-based reporting mechanism only reports when a critical change occurs, whereas the timer-based reporting mechanism reports at every instant when the timer interval has expired even if there is no change between the previously reported data and the next. Additionally, the number of error reversals in the timer-based reporting mechanism is more than that in the threshold-based reporting mechanism, which is true even when τ is small. The reason is that the timer-based reporting mechanism cannot instantly react to the environment changes and thus cause to misuse previously reports severely.

3.2 Packet Loss Effects

To observe how packet loss affects both reporting mechanisms, an actual packet loss distribution is used. Table 2 shows the distribution of the number of consecutive packet loss which is acquired from extensive experiments by transmitting packets at specific moments and recording whether the packets were received or not [10]. The distribution is obtained by the measurement of LoRa packet transmission on NCTU campus (see Fig. 1). A uniform loss of packet is used for comparison. In other words, every time the measured data reported to the IoT server is lost uniformly or based on Table 2.

Figure 5 shows how packet loss affects both reporting mechanisms when uniform-loss and measured-loss distributions are applied in the processes of reporting. As we can see, the results for both the uniform-loss and the measured-loss distributions are about the same ('$*$' and '$\times$' marks, or '$\circ$' and '$\diamond$' markers). With packet loss in the process, one may instinctively think that the degradation of the threshold-based reporting mechanism will be more serious than that of the timer-based reporting mechanism because information of critical changes may be lost more easily. However, Fig. 5 demonstrates that the declined percentages in the timer-based and threshold-based reporting mechanisms are similar.

Table 2. Distribution of the number of consecutive packet losses

# of consecutive packet losses	Frequency	Percentage	Cumulative
1	18,660	84.85	84.85
2	2,596	11.80	96.65
3	496	2.26	98.91
4	123	0.56	99.47
5	30	0.14	99.61
6	22	0.10	99.71
7	18	0.08	99.79
8	8	0.04	99.82
9	7	0.03	99.85
10	4	0.02	99.87
>10	28	0.13	100.00
Total	21,992	100.00	

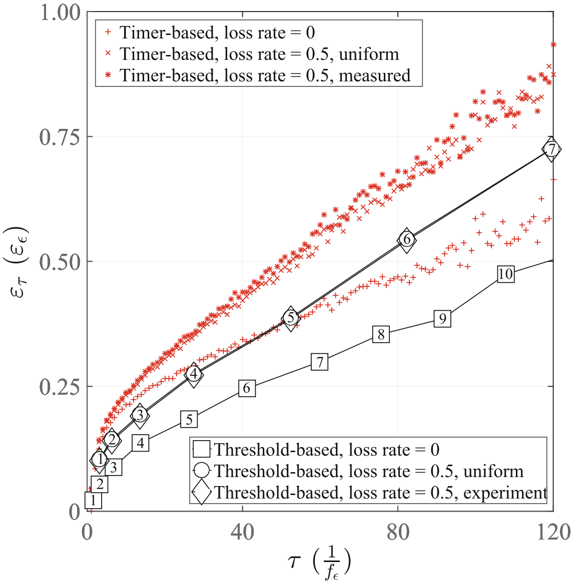


Fig. 5. Expected error ε_τ (ε_ϵ) vs. reporting frequency f_τ (f_ϵ). Note that, the value in a threshold-based marker represents ϵ under particular f_ϵ.

In other words, the threshold-based reporting mechanism is still a better choice even when packets are lost in the communications, which is not trivial[5]. Specifically, the threshold-based reporting outperforms the timer-based reporting by more than 40% in energy saving when the expected errors of both reporting mechanisms are fixed to 0.5. Note that, the expected errors with the loss rate of 0.5 is approximately twice as greater as that with the loss rate of zero, which is a reasonable outcome.

Observation 5: The threshold-based reporting is a better choice than the timer-based reporting with or without packet loss.
Observation 6: Based on the PM2.5 data histograms, the basic strategy to select a τ or an ϵ depends on how much energy to save and how accurate it is required. That is, if ε must be reduced from 0.5 to 0.25, f could be increased from 1/120 to 1/40 by using the threshold-based reporting when there is no packet loss. Additionally, a particular τ or ϵ value that may incur error reversal should not be chosen.
Observation 7: Without from PM2.5 data histogram information, there are several ways to restrain error reversal: (1) using hybrid reporting method by combining the timer-based and threshold-based reporting mechanisms, i.e., reports data either when timer condition or threshold condition is met, (2) irregularly advancing/postponing some reporting points to alter their following reporting points, or (3) cross reporting with multiple settings of τ or ϵ (i.e., interleaved reporting: multiple conditions are used in turn). For example, if two thresholds, ϵ_1 and ϵ_2, are used, a PM2.5 value at t_i will be reported when $|v_i - v_j| > \epsilon_1$ is met. The following PM2.5 value will be reported at t_{i+x} if $|v_{i+x} - v_j| > \epsilon_2$. The reporting behavior will repeat the above two processes one-by-one over and over again [12].

4 Conclusion

In this letter, we investigated two reporting mechanisms to save the energy of IoT devices which are typically installed at hard-to-reach locations on NCTU campus. We discussed how much energy could be saved and how close the reported data to the time-variant environmental information could be. Several interesting findings are provided, including: (1) the threshold-based reporting mechanism is a better choice to report data, (2) reducing reporting frequency may not decrease the data accuracy, and (3) being irregular on reporting at some particular time helps restraining the occurrence of error reversals. In the future, we will apply the adaptive reporting mechanisms to reduce the error reversal phenomenon.

Acknowledgment. This work was supported in part by the Ministry of Science and Technology (MOST) under Grant 106-2221-E-009-006, Grant 106-2221-E-009-049-MY2 and Grant 107-2218-E-009-049, in part by Academia Sinica AS-105-TP-A07, Ministry

[5] Here, we only demonstrate the results where the packet loss probability is set to 0.5. The results with other packet loss probabilities are similar.

of Economic Affairs (MOEA) 106-EC-17-A-24-0619 and the Ministry of Education through the SPROUT Project Center for Open Intelligent Connectivity of National Chiao Tung University and Ministry of Education, Taiwan, R.O.C.

References

1. 3GPP, Introduction of NB-IoT, TS 36.201 (2017)
2. LoRa Alliance. https://www.lora-alliance.org/lorawan-white-papers
3. Sigfox website. http://www.sigfox.com
4. Chang, C.-W., Chen, J.-C.: UM paging: unified M2M paging with optimal DRX cycle. IEEE Trans. Mobile Comput. **16**(3), 886–900 (2017)
5. Chang, C.-W., Chen, J.-C.: Adjustable extended discontinuous reception (eDRX) cycle for idle-state users in LTE-A. IEEE Commun. Lett. **20**(11), 2288–2291 (2016)
6. Liu, B., Yan, S., Li, J., Li, Y.: Forecasting PM2.5 concentration using spatio-temporal extreme learning machine. In: Proceedings of IEEE ICMLA, pp. 950–953 (2016)
7. Tang, M., Wu, X., Agrawal, P., Pongpaichet, S., Jain, R.: Integration of diverse data sources for spatial PM2.5 data interpolation. IEEE Trans. Multimedia **19**(2), 408–417 (2017)
8. Chang,J.-H., Tseng, C.-Y.: Analysis of correlation between secondary PM2.5 and factory pollution sources by using ANN and the correlation coefficient. IEEE Access **5**, 22812–22822 (2017)
9. Kumar, A., Kumar, A., Singh, A.: Energy efficient and low cost air quality sensor for smart buildings. In: Proceedings of IEEE CICT, pp. 1–4 (2017)
10. Reference withheld for double-blind review.
11. Gemtek IoT products web site. http://www.giotnetwork.com
12. Higgins, J.R.: Sampling theory in Fourier and signal analysis: foundations. Oxford Clarendon Press (1996)