Predicting Internet of Things Data Traffic Through LSTM and Autoregressive Spectrum Analysis

Yuhong Li[1], Di Jin[1], Bailin Wang[1]
[1]State Key Laboratory of Networking and Switching Technology
Beijing University of Posts and Telecommunications, Beijing, China

Xiang Su[2,3], Jukka Riekki[3]
[2]Department of Computer Science
University of Helsinki, Helsinki, Finland
[3] Center for Ubiquitous Computing
University of Oulu, Oulu, Finland

Chao Sun[4], Hanyu Wei[4], Hao Wang[4], Lei Han[4]
[4] Network Technology Laboratory
Huawei Technologies Co., Ltd
Nanjing, China

***Abstract*—The rapid increase of Internet of Things (IoT) applications and services has led to massive amounts of heterogeneous data. Hence, we need to re-think how IoT data influences the network. In this paper, we study the characteristics of IoT data traffic in the context of smart cities. Aiming at analyzing the influence of IoT data traffic on the access and core network, we generate various IoT data traffic according to the characteristics of different IoT applications. Based on the analysis of the inherent features of the aggregated IoT data traffic, we propose a Long Short-Term Memory (LSTM) model combined with autoregressive spectrum analysis to predict the IoT data traffic. In this model, the autoregressive spectrum analysis is used to estimate the minimum length of the historical data needed for predicting the traffic in the future, which alleviates LSTM's performance deterioration with the increase of sequence length. A sliding window enables predicting the long-term tendency of IoT data traffic while keeping the inherent features of the data traffic. The evaluation results show that the proposed model converges quickly and can predict the variations of IoT traffic more accurately than other methods and the general LSTM model.**

Keywords—Internet of Things; data traffic prediction; LSTM; autoregressive spectrum analysis

I. INTRODUCTION

The massive amount of data from Internet of Things (IoT) has a significant influence on network infrastructure and applications running on it. Therefore, it is important to predict the IoT data traffic at different time lags and intervals in order to provide networking performance guarantees for both IoT and other applications and to optimize network resources. However, the amount of data generated by IoT applications varies greatly, from several bits generated by simple sensors to several hundred megabytes generated by video cameras. The pattern of data generation also varies from periodic to sporadic. The same application produces different amount of data on different platforms. Thus, the special characteristics of IoT data make it difficult to predict IoT traffic and estimate its influence on public networks and infrastructure facilities. Especially, different platforms and network architecture are used in different IoT applications and scenarios [1]. How these platforms influence network traffic is still unclear.

Widely accepted methods for modelling and predicting traffic are based on the Internet [2][3]. They cannot be used straightforwardly in the IoT environment due to the different behaviour of IoT devices and human beings using the Internet. Traffic modelling has been studied in the context of machine-to-machine (M2M) and ubiquitous sensor networks (USN) [4]-[7]. However, only the traffic model of a single application type has been considered, such as telemetry [8], video surveillance [9], and YouTube [10]. Although modelling and prediction of the aggregated IoT traffic from several applications have been studied [11][12] in the context of building smart cities and smart homes recently, these work are all based on mathematical models. Several data-driven models, such as deep learning [13][14] have been proposed in the context of smart transportation. However, the focus has been on predicting vehicular traffic congestion, which cannot be used straightforwardly for predicting IoT data traffic due to the different features of the time sequences of IoT data and vehicular traffic. For example, vehicular traffic might have strong dependency on the weekday or the time of a day. In this paper, we generate IoT data by simulating different types of IoT applications according to the models revealed by other researchers [15]-[20], and propose a data-driven model to predict the aggregated IoT data traffic.

Regarding data-driven models for time-sequence prediction, methods based on Autoregressive Integrated Moving Average Model (ARIMA ARMA/ARIMA) [21] and the HoltWinters algorithm [22] are the most widely used linear prediction methods. Since they cannot capture the nonlinear features of data traffic, in general they are only suitable for predicting the time sequences whose underlying features are dominated by linear functions. Methods based on recurrent neural network (RNN) are non-linear methods. Since they can learn from experience and can capture both the linear and nonlinear features in a dataset, they have been used for data prediction. Particularly the Long-Short Term Memory (LSTM) algorithm, due to its long and short term memory capabilities, has been used widely for predicting network traffic matrix [23] and time sequences [24]. However, most of the work has focused on real-time prediction, where continuous feeding and learning are used. Especially, the performance of LSTM deteriorates rapidly as the length of the input sequence increases [25]. This makes it difficult to use LSTM to predict IoT data traffic, since long sequences of historical data are needed because of the long range dependency of IoT data traffic, as shown in Table IX. Recently, mechanisms have been studied in order to improve LSTM's lack of learning ability for extremely long sequences [14]. However, this work is based on the assumption that the vehicular traffic volume of the previous day has big impact on the forecast. Our work aims at alleviating LSTM's long-term dependency problem and

978-1-7281-4973-8/20/$31.00 ©2020 IEEE

keeping the prediction model as simple as possible, so that it can be used for relatively long-term prediction to realize network and resource optimization by network operators.

There are also several hybrid models for predicting time sequences [26][27]. These models decompose the time sequences into linear and non-linear, or lowly and highly volatile components, or add new layers to LSTM. However, if the linear component of the time sequences cannot be well recognized and predicted, as is the case with aggregated IoT traffic, these models do not produce accurate prediction results. Our model extracts the basic features from the given long data values dynamically and feeds them to LSTM. There is no fundamental separation between the linear and non-linear components. Therefore, neither the linear nor the non-linear prediction errors will be accumulated, and prediction accuracy can be kept without using long sequences of historical data.

The main contributions of our paper are threefold.

- The aggregated IoT data traffic at different network levels is simulated and collected according to the research results of other people, and self-similarity and long-range dependency characteristics of the traffic are analyzed.
- A data-driven model is proposed to predict the aggregated IoT data traffic. AR spectrum analysis is used to capture the basic features of the data traffic, estimating the minimum length of the given data needed for predicting future traffic dynamically, and LSTM is used to predict the future traffic values. This alleviates the problem of LSTM's performance deterioration with the increase of sequence length with simple linear computations.
- The performance of the proposed model is evaluated. The predicted value and the actual value, as well as the Mean Absolute Percentage Error (MAPE) in different situations are compared, which demonstrates that the proposed method can well predict the IoT data traffic tendency in a simpler way.

The remainder of the paper is organized as follows: we elaborate the simulation of the IoT data traffic and analyze the characteristics of the aggregated data in Section II. Then we describe the proposed model and give the evaluation results in Section III and IV, respectively. We discuss future research and conclude the paper in Section V.

II. IoT Traffic Simulation and Analysis

A. IoT Applications and Traffic

IoT traffic is unique due to its independency from human behaviour. In particular, the time intervals of the information exchange are not defined by humans. Since the Smart City use case includes most of the applications whose data traffic influences greatly the network infrastructure, we focus on this use case. Here, the traffic may include four types of data, namely heartbeat, event trigger, payload data, and node update. Among them, the heartbeat type is generally triggered by devices themselves, and used for exchanging the status information of the nodes for staying in the network. The traffic pattern is regular with constant size of packets. The event trigger type is usually initiated by operators, servers or other devices in order to trigger an action of a device, such as reporting the temperature or turning on or off lights. Trigger processes are irregularly generated by the servers or other devices, and the data can be sent in very short packets. Contrast to this, the payload data type carries the information exchanged between devices and servers through networks. There are no strict rules for this type of traffic. Its size can be constant, e.g. telemetry, or variable, e.g. images in video surveillance. The node update type is used to maintain the devices, e.g., a server in a network may push new configurations or firmware to devices. Updates may not happen very often, but might produce bursts to the network.

Consisting of different types of data, the aggregated data traffic generated by large numbers of heterogeneous IoT devices has the following characteristics:

- The aggregated traffic is heavily dependent on each type of application scenario. In the same application scenario, the traffic varies also with the number of IoT devices.
- The traffic pattern of a single device and a single scenario might be simple. However, when the amount of scenarios and devices sharing a network increases, the aggregated traffic pattern becomes complicated. This may result in an unknown influence on the network infrastructure.
- Since new applications and scenarios may emerge frequently and new networking technologies may be used to provide better services to IoT applications, the total IoT traffic may change dynamically in an unexpected way.

Hence, it is very important to predict the aggregated IoT data traffic using a data-driven method. However, the restrictions set by network operators prevent collecting IoT data from real networks. To the best of our knowledge, there is no public IoT dataset generated by a large amount of heterogeneous devices. Therefore, we first generate a large set of IoT data traffic by simulating different applications according to the data patterns revealed by other researchers, and then evaluate our model using the generated dataset.

B. IoT Data Traffic Simulation

Fig. 1 illustrates the network configuration we use in the network simulator NS-3 [28], which is based on the Smart City use case [29], where various applications with different traffic patterns are involved. To obtain the aggregated traffic at different levels, we design eight subnetworks (we refer each of them also as a scenario), each with a gateway (we refer a gateway at the subnetwork level as a sink). The traffic from each sink is aggregated at the gateways (i.e., 1-3) connecting to the Internet through Gateway 4. To obtain as realistic IoT data as possible, we generate the data of each type of applications according to the characteristics revealed by other researchers. In addition, we set multiple applications in each scenario, e.g., 800 applications (i.e., machines) for goods tracking are set in scenario (a). Through the configurations described below, we obtain the traffic aggregated from a large amount of devices.

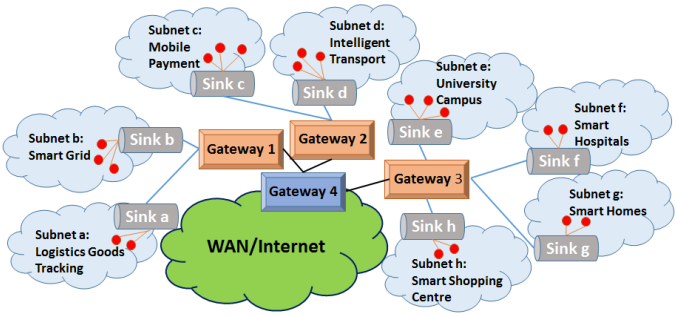


Fig. 1. Network configuration for IoT data traffic simulation.

(a) Logistics Goods Tracking [5]. The data sent from the devices for logistics goods tracking follows the On-Off model, where the On and Off time follows Pareto distribution. Packet intervals during On time follow the normal distribution with parameters of (0.2, 1) (Table I). Data from 800 goods tracking are simulated.

TABLE I. TRACK PATTERN OF LOGISTICS GOODS TRACKING

	Pareto distribution	*Packet size*	*Packet intervals*
On	(0.60, 19.654,4.99)	50 bytes	Normal distr.(0.2, 1)
Off	(0.009,10.026,0.001)	N.A.	N.A.

(b) Smart Grid [7]. The data traffic caused by one voltage transformer follows On- Off model (Table II). In total, data from 6000 transformers are collected.

TABLE II. TRACK PATTERN CAUSED BY ONE VOLTAGE TRANSFORMER

	Time (min)	*Packet size(byte)*	*Packet intervals*
On	5	135	Exp. distr. (μ=60s)
Off	15	N.A.	N.A.

(c) Mobile Payment [15]. The payments from each machine follows exponential distribution (Table III). Data from 2050 payment machines are collected.

TABLE III. TRAFFIC PATTERN OF EACH MOBILE PAYMENT

	Packet size (byte)	*Packet intervals*
Data type	205	Exp. distribution (μ=2.5s)

(d) Intelligent Transport [16]. The data traffic produced by each vehicle is illustrated in Table IV. Each vehicle produces data at an interval following the normal distribution. Altogether data produced by 2000 vehicles with different starting times are collected.

TABLE IV. TRAFFIC PATTERN OF EACH VEHICLE

	Packet size (byte)	*Packet intervals*
Data type	149	Normal distribution (80,8)

(e) University Campus [17]. Among the five types of sensors generally utilized in campuses (Table V), three types generate data with fixed length periodically. The surveillance video cameras generate data periodically but the packet lengths follow the lognormal distribution. The intervals of check-in service follow normal distribution. Altogether data from 15 universities are simulated.

TABLE V. IOT DEVICES AND TRAFFIC PATTERN OF EACH UNIVERSITY

Device type	*Device No.*	*Packet size (byte)*	*Period (s)/ Intervals*
Humidity detector	50	90	1800
Temperature sensor	50	80	600
Smoke detector	60	80	600
Surveillance video (security)	45	Lognormal distr.(5.9,1.2)	15
e-card check-in	1050	205	Normal distr.(20,2)

(f) Smart Hospital [18]. *The* sensors and traffic patterns of equipment in one hospital are illustrated in Table VI. There are 50 pieces of equipment in each hospital, and data from 10 hospitals are generated.

TABLE VI. SENSORS AND TRAFFIC PATTERN IN EACH HOSPITAL

Device type	*Packet size (byte)*	*Period (s)*
cardiograph	125	0.2
Blood pressure	125	0.5
Pulse	125	0.2
Body temperature	27	5

(g) Smart Homes [19]. The traffic patterns of each household are shown in Table VII. Altogether data from 100000 households are generated.

TABLE VII. IOT DEVICES AND TRAFFIC PATTERN OF EACH HOUSEHOLD

Device type	*Device No.*	*Packet size (byte)*	*Period (s)*
Humidity detector	5	90	1800
Temperature sensor	4	80	600
Smoke detector	2	80	600
Light switch	5	100	1800
Sensor for sec. system	1	80	600
Electric meter	1	90	600

(h) Smart Shopping Centre [20]. The data intervals of payment and retail system follow exponential and Pareto distribution respectively (Table VIII). Data of 8 shopping centres are simulated.

TABLE VIII. DEVICES AND TRAFFIC PATTERN OF EACH SHOPPING CENTRE

Device type	*Device No.*	*Packet size (byte)/distr.*	*Period (s)/ Intervals*
Humidity detector	20	90	1800
Temperature sensor	10	80	600
Smoke detector	25	80	600
Surveillance (security)	20	Log normal distr.(5.9,1.2)	15
Payment	10	205	Exp. distr. (μ=2.5 sec.)
Retail system	10	72	Pareto distr.(1,10)

Since in practice all devices do not start sending data simultaneously, we assume:

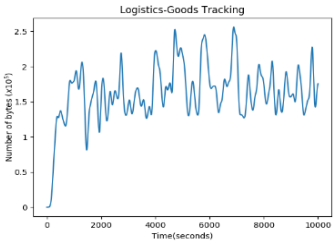
(1) For one scenario (e.g., Logistics Goods Tracking or University Campus), all the devices or sensors are randomly started within 100 seconds when the whole simulation time is 10000 seconds. For example, for the logistics goods tracking, each of the 800 devices starts with On state, but after a random delay within 100 seconds. For the University Campus, the data traffic aggregated in one university will be added with another after a random delay within 100 seconds.

(2) Within one scenario, the devices of the same type start sending data following Poisson (30, 50) distribution. Data from different types of devices are aggregated together, which forms the data of one "region" for that scenario. For example, for

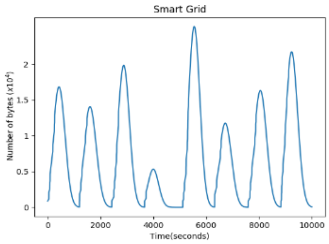
each campus, 50 humidity detectors start to send data following the Poisson distribution. All the data from different types of devices consist of the data of one university campus.

C. Simulation Results and Analysis

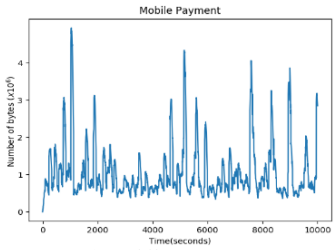
Fig. 2 illustrates the collected data during a period of 10000 seconds.



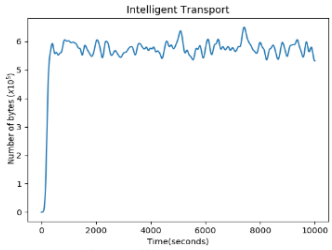
(a) Logistics Goods Tracking



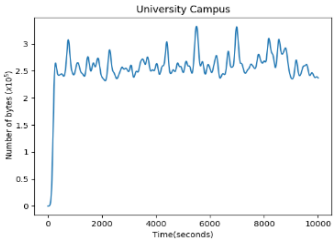
(b) Smart Grid



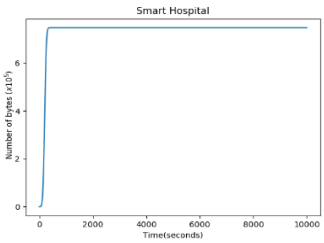
(c) Mobile Payment



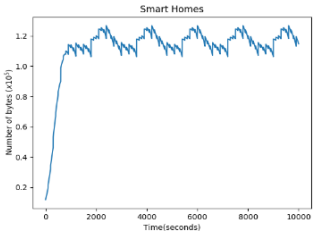
(d) Intelligent Transport



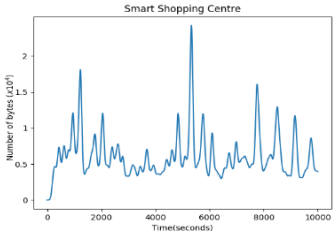
(e) University Campus



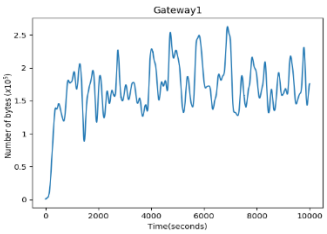
(f) Smart Hospital



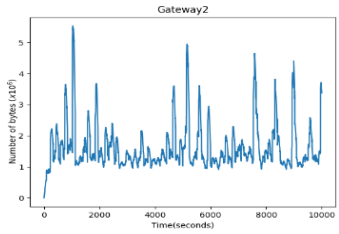
(g) Smart Homes



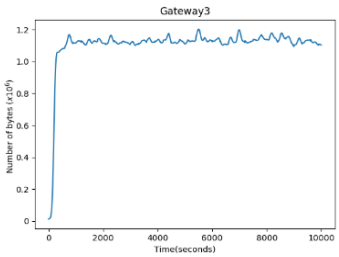
(h) Smart Shopping Centre



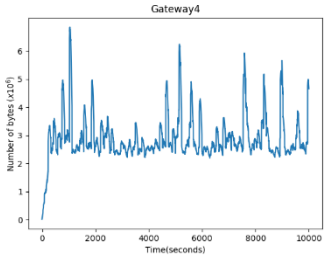
(i) Gateway1



(j) Gateway 2



(k) Gateway 3



(l) Gateway 4

Fig. 2. Aggregated data at sinks and gateways for 10000 seconds.

From Fig.2, we can observe that the aggregated data do not have any particular patterns and may have bursts, although the data sent by each device of an application scenario follows certain distribution and can be well modelled and predicted.

In order to investigate the stationary and autocorrelation of the data, we performed the Augmented Dickey–Fuller (ADF) and Ljung-Box tests on the data. The intuition behind ADF test is that if the series is integrated, the lagged level of the series y_{t-1} will provide no relevant information in predicting the change in y_t besides the one obtained in the lagged changes Δy_{t-k}. Ljung-Box test is performed to see whether any group of autocorrelations of a time series are different from zero. The results are shown in Table IX. Moreover, to estimate the long-range dependence (LRD) of the IoT traffic, we also calculate the Hurst exponent of the IoT traffic, which is referred to "the index of dependence" or "the index of long-range dependence". As shown in Table IX, the value H at all the gateways are bigger than 0.5, and all the p-value are very small, which means the IoT traffic is a stationary LRD time series.

TABLE IX. ADF AND LJUNG-BOX TEST RESULTS

	Box-Ljung	***ADF***	***Hurst exponent H***
Gateway1	p-value < 2.2e-16	stationary	0.8233349
Gateway2	p-value < 2.2e-16	stationary	0.7289894
Gateway3	p-value < 2.2e-16	stationary	0.7995602
Gateway4	p-value < 2.2e-16	stationary	0.7344491

III. LSTM MODEL ENHANCED WITH AR SPECTRUM ANALYSIS

A. Insights of Long Short-Term Memory and Autoregressive Spectrum Analysis

LSTM [30] mitigates the vanishing gradient problem of traditional RNN, which makes LSTM suitable to model the long-term dependencies of IoT traffic. The LSTM architecture consists of an input layer, a recurrent hidden layer whose basic unit is a memory block (instead of a traditional neuron node), and an output layer. Memory blocks are a set of recurrently connected subnets. Each block contains one or more inter-connected memory cells and three multiplicative units - the input, output and forget gates, which provide the continuous analogues of write, read and reset operations on the cells. The multiplicative gates allow LSTM memory cells to store and access information over long periods of time, which is essential for modelling time sequences that is LRD.

However, due to the chain structure of the memory blocks, the performance of LSTM deteriorates rapidly as the length of the sequence increases. Yet the aggregated IoT data traffic is

long-range dependent (i.e., autocorrelated) - the IoT data in the long past will influence the traffic in the future. Therefore, finding the proper time lag, that is, how long historical traffic sequence is needed to predict future traffic, is the key to reduce the complexity (i.e., long-term dependency) of LSTM and at the same time to maintain prediction accuracy.

We approach this challenge by augmenting LSTM with spectrum analysis. That is, our aim is not to improve LSTM itself. Spectrum analysis can be used to estimate the energy spectrum of a stationary stochastic process by a given number of samples of the process. AR spectrum analysis is one of the widely used techniques for spectrum analysis, as it needs only to solve the linear equations which can significantly reduce computations when compared with methods used for solving nonlinear equations.

We use the AR spectrum analysis to find out the length of the historical data (i.e., the order of the time sequence) that influences the prediction accuracy of future values. This approach is based on the consideration that for stationary time sequences, energy spectrums are roughly constant. Thus, the length of the historical sequence that needs to be considered can be identified based on the spectrum of the time sequence. Moreover, the linear computations of AR analysis can decrease the computation complexity of the model greatly, which helps to realize the relative long-term prediction.

B. LSTM Model enhanced with AR Spectrum Analysis

Based on these considerations, our model first uses the AR spectrum analysis to estimate how long data sequence from the past will mostly affect the traffic in the future for a given time series, then uses LSTM to train the neural network by using the selected length of the traffic and predict the traffic in the future. Fig. 3 illustrates the model.

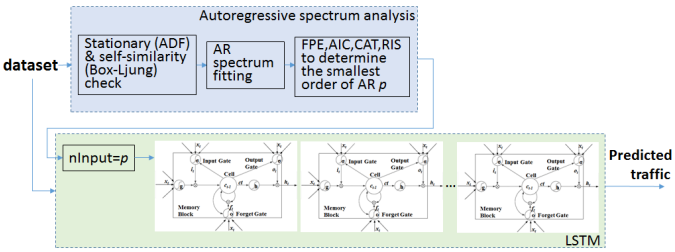


Fig. 3. LSTM model based on AR spectrum analysis.

The ADF and Box-Ljung tests are performed before the AR spectrum fitting, in order to check the stationary and self-similarity (i.e., long-term dependence) of the IoT data traffic.

The AR model of time series $x(n)$ can be defined as

$$x(n) = -\sum_{k=1}^{p} a_k x(n-k) + w(n)$$

where $w(n)$ is the white noise which has the mean of zero and a variance of σ_w^2 , p is the order of the AR model, and a_i is the coefficients of the model.

The Z-transformation $H(z)$ of $x(n)$ is:

$$H(z) = \frac{X(z)}{W(z)} = \frac{1}{1+\sum_{k=1}^{p} a_k z^{-k}}$$

Thus, the power spectrum of can be obtained as:

$$P_x(w) = \sigma_w^2 \left|H(e^{jw}\right|^2 = \frac{\sigma_w^2}{\left|1+\sum_{k=1}^{p} a_k e^{-jwk}\right|^2}$$

We use Levinson-Durbin algorithm to perform regression during the spectrum fitting, thus the autoregressive spectrum of $x(n)$ can be estimated directly from a_k without generating $x(n)$.

We use four criteria to determine the optimum order p of the AR:

- Akaike's FPE (Final Prediction Error):

$$FPE = \sigma_P^2 \left(\frac{N+p+1}{N-p-1}\right)$$

- AIC (Akaike Information Criterion)

$$AIC = In(\sigma_P^2) + \frac{2(p+1)}{N}$$

- Parzen CAT (Criterion Autoregressive Transfer Function)

$$CAT = \left(\frac{1}{N}\sum_{j=1}^{p} \frac{N-j}{N\sigma_j^2}\right) - \frac{1}{\sigma_p^2}$$

- Rissanen (RIS) minimum description length, namely

$$RIS = \sigma_p^2 \left(1 + \left(\frac{p+1}{N}\right) In(N)\right)$$

By using the Burg algorithm, the optimum order of AR can be selected.

During the training phase, p values of the collected traffic were used as a block one by one to train the parameters of the LSTM. There are two common algorithms to train the RNN - Back Propagation Through Time (BPTT) and Real Time Recurrent Learning (RTRL) [31]. We use BPTT algorithm in our study due to its simplicity and efficiency in computation time.

After training, we use m-value sliding window to predict the time series, as shown in Fig. 4, in order to realize relatively long term prediction.

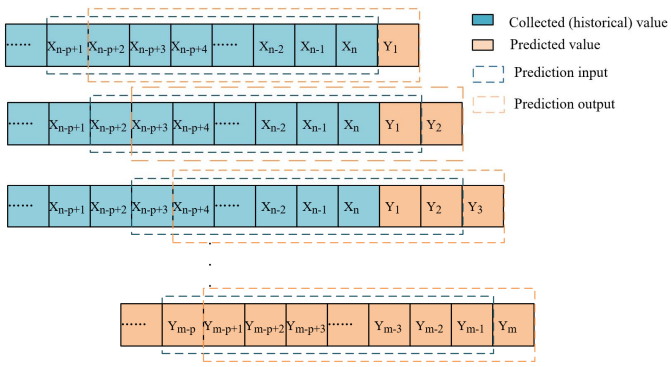


Fig. 4. m-value sliding prediction.

Suppose the length of the collected traffic is n and the number of m values of the traffic will be predicted. In this case, only the last p of the n values of the obtained traffic series are selected during the prediction phase. After p times of prediction, the predicated values are used to further predict future values.

After m times of iterated prediction, all the needed values can be predicted.

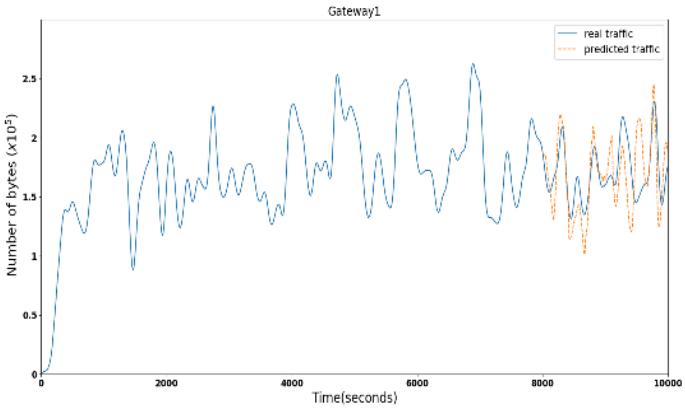
IV. Prediction Results and Analysis

To evaluate the proposed model, we use 8000 values of the collected data given in Fig. 2 as the datasets to perform the AR spectrum analysis and to train the LSTM model, and to predict 2000 values of the future, i.e., m=2000 in Fig.4, which is relative long-term prediction given the historical values. Then we compare the predicted value with the remaining 2000 values of the collected data (i.e., as the real observed value) to show the accuracy of our proposed model. To show the advantages of introducing AR spectrum analysis, we train and predict the LSTM with different input length too. During the evaluation, the commonly used Mean Absolute Percentage Error (MAPE) is used, which is defined as follows:

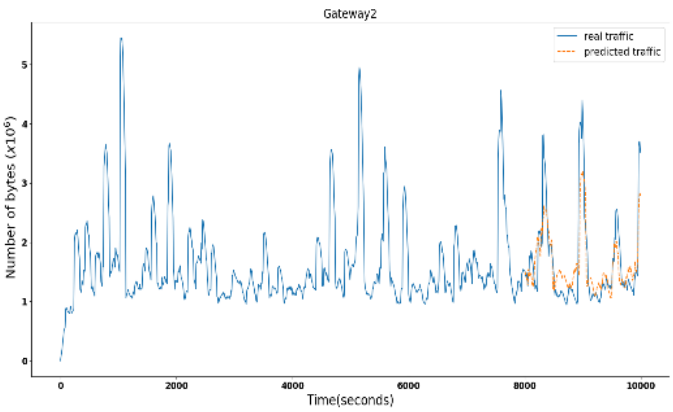
$$MAPE(v_1, v_2) = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{v_1 - v_2}{v_1}\right|$$

where v_1 and v_2 are the observed and predicted value respectively.

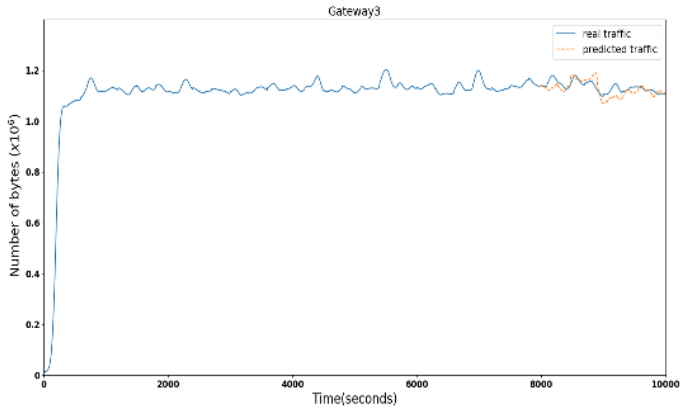
Fig. 5(a)-5(d) illustrates the comparison of the observed and the predicted aggregated IoT traffic at Gateway 1 to 4 shown in Fig. 1. The corresponding MAPE of Fig. 5(a)- 5(d) are 0.1211, 0.1680, 0.0158, 0.1243, 0.0613 respectively. Here we can see that our model can well locate the burst of the traffic, and predict the tendency of the traffic.



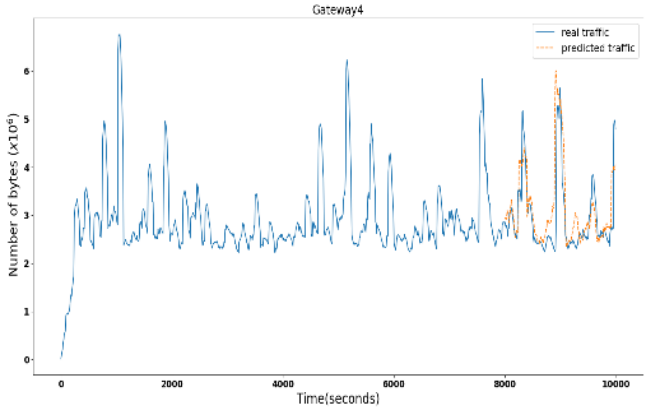
(a) Gateway 1



(b) Gateway 2



(c) Gateway 3



(d) Gateway 4

Fig. 5. Comparison of observed and predicated traffic.

To compare the results with short-term prediction, we set m=1, i.e., in each recurrent step, only one value is predicted. Fig. 6 shows the results at Gateway 1. The MAPE of the short-term and long-term prediction is 0.0141 and 0.1211 respectively. In other words, long-term prediction is obtained at the cost of accuracy.

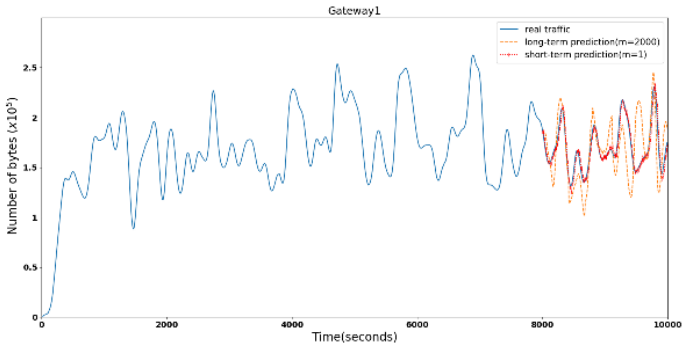
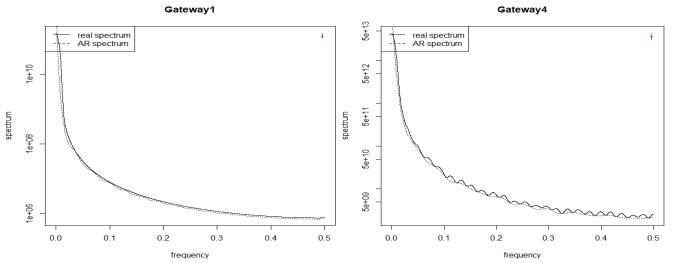


Fig. 6. Comparison of short and long term prediction.

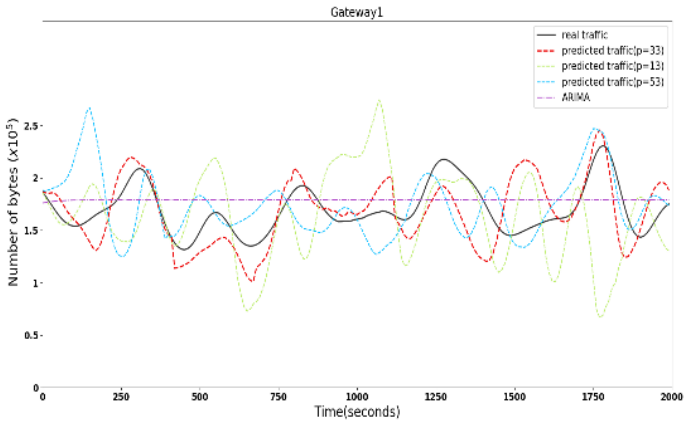
Fig. 7 illustrates the spectrum of real traffic (i.e., collected traffic) and the estimated AR spectrum for the LSTM training and prediction. Due to the space limitation, only the spectrums at Gateway 1 and 4 are shown.



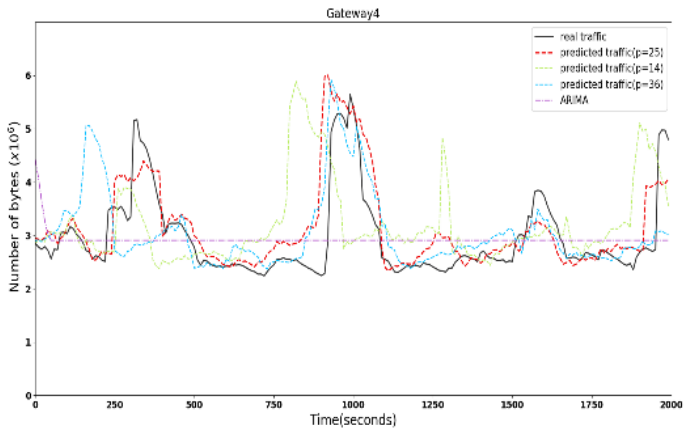
(a) Gateway 1 (b) Gateway 4

Fig. 7. AR spectrum of Gateway 1 and 4.

We performed also predictions using different lengths of historical data, i.e., p=13 and 53 for Gateway 1, p=14 and 36 for Gateway 4, in order to compare them with the prediction results of using AR spectrum analysis, i.e., p=33 for Gateway 1, and p=25 for Gateway 4. We also illustrate the results of using the linear prediction method ARIMA [21]. Fig. 8 shows the results of the real traffic and the predicted traffic. Due to the space limitation, we only illustrate the results of Gateway 1 and 4.



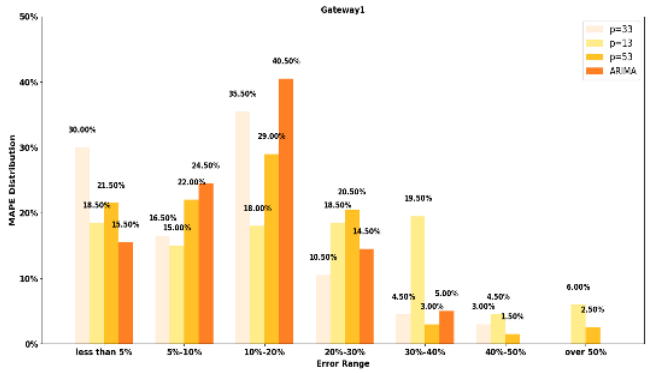
(a) Gateway 1



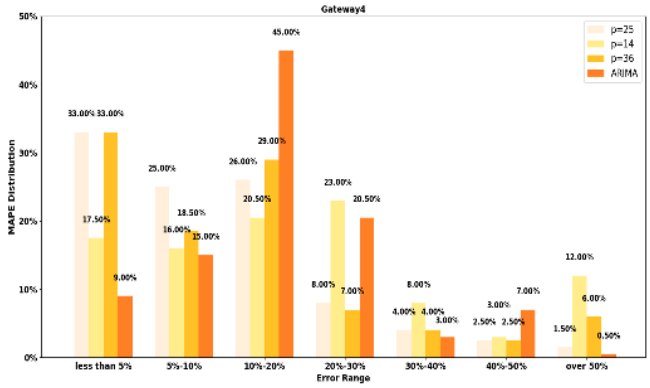
(b) Gateway 4

Fig. 8. Predictions with different historical data lengths in LSTM and ARIMA.

Fig. 9 shows the MAPE distribution. Here we can see that the accuracy of the prediction result does not increase with the increase of the historical length used in LSTM. When the historical length p is equal to 33 for Gateway 1 and 25 for Gateway 4, as calculated by using the AR algorithm, the MAPE achieves the smallest value. This is because AR can find the average period of long-term dependency hidden in the given data. By using the selected historical traffic sequence, the accuracy of the LSTM prediction can be improved.



(a) Gateway 1



(b) Gateway 4

Fig. 9. MAPE distribution of LSTM and ARIMA.

Table X shows the comparison of MAPE and the training time by using different p at Gateway 1 and Gateway 4. Here we can see that selecting higher p cannot obtain better MAPE but the training time will increase remarkably.

TABLE X. COMPARISON OF MAPE AND TRAINING TIME

	MAPE			Training time (minutes)		
	smaller p	***p***	***bigger p***	***smaller p***	***p***	***bigger p***
Gateway 1	0.2119	0.1211	0.1407	61	137.5	216.5
Gateway 4	0.2535	0.1243	0.1455	68	113	150

V. CONCLUSIONS AND FUTURE WORK

In this paper, we presented the results of predicting IoT data traffic by combining LSTM with AR spectrum analysis. We studied the characteristics of IoT data traffic in the context of the Smart City. We designed a network architecture with hierarchical gateways and generated the data traffic aggregated at different levels of gateways to simulate the IoT traffic in different real-world applications. We analyzed the

characteristics of the traffic. We proposed the LSTM model combined with AR spectrum analysis to predict the aggregated traffic from different IoT scenarios. The proposed model uses AR spectrum analysis to extract the feature of the data traffic, estimating the minimum length of the historical data needed for predicting the traffic in the future, then uses LSTM to predict the traffic in the future. The evaluation results show that the proposed model can well predict the relatively long-term tendency of the traffic considering the training time, computing complexity together with the prediction accuracy.

As future work, we will analyze the factors affecting the prediction accuracy and refine the proposed model. In addition, we will evaluate the model with more datasets and more configuration parameters of networks. Moreover, we will evaluate the proposed model by using the IoT data collected from public networks, e.g., 4G/5G base stations. We will also investigate the impact of new network technologies, such as 5G, to the IoT applications.

REFERENCES

[1] Y. Li, X. Su, J. Riekki, T. Kanter, R. Rahmani, "A SDN-based architecture for horizontal Internet of things services," In: Prof. of IEEE International Conference on Communications (ICC), Kuala Lumper, pp. 1-7, 2016.

[2] W. Leland, M. Taqqu, W. Willinger, D. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," IEEE/ACM Transactions on Networking, vol. 2, no. 1, pp. 1-15, 1994.

[3] A. Feldmann, A.C. Gilbert, W. Willinger, "Data networks as cascades: Investigating the multifractal nature of internet WAN traffic," In: Proc. of SIGCOMM, pp. 42-55, 1998.

[4] O. Al-Khatib, W. Hardjawana, B. Vucetic, "Traffic modeling for machine-to-machine (M2M) last mile wireless access networks," In Proc. of Communications QoS, Reliability and Modelling Symposium, Austin, TX, USA, pp. 1199-1204, 2014.

[5] Q. Wang and T. Zhang, "Source traffic modelling in wireless sensor networks for target tracking," In Proc. of the 5th ACM International Symposium on Performance Evaluation of Wireless Ad hoc, Sensor and Ubiquitous Networks, Vancouver, Canada, Oct. 27-31, pp. 96-100, 2008.

[6] F. Al-Turjman, A. Radwan, S. Mumtaz, J. Rodriguez, "Mobile traffic modelling for wireless multimedia sensor networks in IoT," Computer Communications, vol. 112, pp. 109-115, November 2017.

[7] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, J. Wang. "Large scale measurement and characterization of cellular machine-to-machine traffic," IEEE/ACM Transactions on Networking (ToN), vol. 21, no. 6, pp. 1960-1973, December 2013.

[8] A. Koucheryavy, A. Prokopiev, "Ubiquitous sensor networks traffic models for telemetry applications," In: Proc. of NEW2AN 2011, vol. 6869, pp. 287-294, 2011.

[9] I. Petiz, P. Salvador, A. Nogueira, "Characterization and modeling of M2M video surveillance traffic," In Proc. of the Fourth International Conference on Advances in Future Internet (AFIN 2012), Rome, Italy. pp. 65-70, 2012.

[10] G. Horváth, P. Fazekas, "Characteriztion and modelling of YouTube traffic in mobile networks," In: Proc. of the Fourteenth International Conference on Networks (ICN 2015), pp. 115-121, 2015.

[11] S. Wu , W. Mao, C. Liu , and T. Tang, "Dynamic traffic prediction with adaptive sampling for 5G HetNet IoT applications", Wireless Communications and Mobile Computing, Wiley, Vol. 2019, Article ID 4687272.

[12] C. Majumdar, M. López-Benítez and S. N. Merchant, "Accurate modelling of IoT data traffic based on weighted sum of distributions", In: Prof. of IEEE International Conference on Communications (ICC), Shanghai, China, 2019.

[13] Z. Abbas, A. Al-Shishtawy, S. Girdzijauskas, V. Vlassov, "Short-term traffic prediction using long short-term memory neural networks," In: Proc. of IEEE International Congress on Big Data, July 2-7, 2018. San Francisco, USA, pp. 57-65, 2018.

[14] B. Yang, S. Sun, J. Li, X. Lin, Y. Tian, "Traffic prediction using LSTM with feature enhancement," Neurocomputing, vol. 332, pp. 320-327, March 2019.

[15] C. Zhang, H. Wang, Y. Cui, J. Wu, W. Wang, "TSWiFi: An optimal payment-based traffic sharing in mobile networks," In: Proc. of IEEE International Conference on Broadband and Photonics, Bali, pp. 1-6, 2015.

[16] D. Drajic, S. Krco, et al., "Traffic generation application for simulating online games and M2M applications via wireless networks," In: Proc. of 9th Annual Conference on Wireless On-demand Network Systems and Services (WONS), Courmayeur, Italy, pp. 167-174, 2012.

[17] Y. Luo, X. Guo, "Research of pedestrian traffic characteristics in university campus," In: Proc. of The Fourth International Conference on Digital Manufacturing and Automation (ICDMA), Qingdao, pp. 1022-1024, 2013.

[18] A. Ahmad, A. Riedl, W. J. Naramore, N. Chou, M. S. Alley, "Scenario-based traffic modeling for data emanating from medical instruments in clinical environment," In: Proc. of 2009 WRI World Congress on Computer Science and Information Engineering, Los Angeles, CA, pp. 529-533, 2009.

[19] A. Orrevad, "M2M traffic characteristics: When machines participate in communication," Master of Science Thesis, Royal Institute of Technology, Sweden, 2009.

[20] S. B. Osoba, "Appraisal of parking problems and traffic management measures in central business district in Lagos, Nigeria," Journal of Sustainable Development, vol. 5, no. 8, pp. 105-115, 2012.

[21] I. Rojas, O. Valenzuela, et al., "Soft-computing techniques and ARMA model for time series prediction," Neurocomputing vol. 71, no. 4-6, pp. 519-537, January 2008.

[22] J. G. De Gooijer, K. Kumar, "Some recent developments in non-linear time series modelling, testing, and forecasting," International Journal of Forecasting, vol. 8, no. 2, pp. 135-156, October 1992.

[23] A. Azzouni, G. Pujolle, "A long short-term recurrent neural network framework for network traffic matrix prediction," arXiv:1705.05690v3, hal-01538471, 2017.

[24] Y. Qin , D. Song , H. Chen , et al. , A dual-stage attention-based recurrent neural network for time series prediction, IJCAI (2017) 2627–2633.

[25] K. Cho , B. Van Merriënboer , D. Bahdanau , et al. , "On the properties of neural machine translation: Encoder-decoder approaches", Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, pp. 103–111, 2014.

[26] C. N. Babu, B. E. Reddy, "A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data," Applied Soft Computing, vol. 23, pp. 27-38, October 2014.

[27] Y. Xiao, Y. Yin, "Hybrid LSTM neural network for short-term traffic flow prediction," Information, vol. 10, no. 3, March 2019.

[28] NS-3: A Discrete-event Network Simulator. http://www.nsnam.org/

[29] White paper, Standardization of Chinese Smart City, 2014. (in Chinese).

[30] S. Hochreiter, J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[31] F. J. Chang, L. C. Chang, H. L. Huang, "Real-time recurrent learning neural network for stream-flow forecasting," Hydrological Processes, vol. 16, no. 13, pp. 2577-2588, 2002.