A Two-Level Hidden Markov Model for Characterizing Data Traffic from Vehicles

Yuhong Li, Xiaoyu Hao, Han Zheng
State Key Laboratory of Networking and Switching Technology
Beijing University of Posts and Telecommunications
Beijing, China
hoyli@bupt.edu.cn

Xiang Su, Jukka Riekki
Centre for Ubiquitous Computing
University of Oulu
Oulu, Finland
xiang.su@oulu.fi,
jukka.riekki@oulu.fi

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

ABSTRACT
With the popularization of intelligent transport and mobile internet services, vehicles and people on board generate increasing amounts of data. To match future networks with this use case, tools are needed to analyze the requirements set for the network. In this paper, we study the characteristics of data traffic in the context of networked vehicles. We generate data traffic based on real-world vehicle traces and reported data patterns of end-user applications and vehicles. Based on this, we propose a two-level hidden Markov model to describe both large and small temporal characteristics of data traffic from vehicles aggregated on base stations. We evaluate the proposed model by comparing the original and synthesized data. The results show that the proposed model can well characterize the data traffic from vehicles.

Author Keywords
Traffic modelling; Hidden Markov Model; Vehicular traffic; Internet of Things.

ACM Classification Keywords
• Networks → Network performance evaluation → Network performance modelling

INTRODUCTION
With the popularization of intelligent transportation and mobile internet services, large amounts of data from vehicles and people on board need to be sent through networks. The services used by vehicle drivers and passengers generate data. Intelligent transportation requires monitoring vehicle behavior and environment states and thus exchanging information through networks. Furthermore, vehicles are regarded as useful carriers for crowd-sourcing, collecting data about road condition, traffic density, unmanned vehicles etc. As a result, large amounts of data need to be transmitted, often to the cloud. The data traffic places a burden specifically on the network infrastructure near the roads.

Predicting the amount of data traffic generated in this use case would facilitate optimizing the network performance and designing future networks. However, modelling such data traffic is challenging. As the vehicles are in motion and the amount of vehicles and the behavior of the people vary during the day, the network load changes dynamically and is heavily unbalanced. Due to this dynamicity and the complexity of the generated data, the widely accepted traffic models for Internet [1-3] cannot be directly used in this use case. The recent traffic models for machine-to-machine traffic [4][5] and video streaming traffic [6] in cellular networks do not take into account the data traffic aggregated in this use case. More analysis about the traffic modelling methods see the section below. Moreover, the restrictions set by network operators prevent building a model based on data traffic monitored from real vehicles. Obviously, this is the case with future scenarios as well. Hence, new solutions are needed.

In this paper, we use the available information to build a model for the data traffic produced by vehicles and people on board. We simulate data traffic and analyze the characteristics of different types of data produced by both the vehicular devices, such as sensors, and the devices used by drivers and passengers, taking the behavior of people into account. We propose a two-level hidden Markov model to characterize the data traffic aggregated on the base stations of mobile cellular networks.

The rest of the paper is organized as follows. We first describe the data traffic from vehicles and discuss related work about data traffic modelling. Then we introduce the method for obtaining the data traffic used for our modelling. Following this, we describe the proposed modelling method. Finally, we present the evaluation and summarize the results.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
IoT 2017, October 22–25, 2017, Linz, Austria
© 2017 Association for Computing Machinery.
ACM ISBN 978-1-4503-5318-2/17/10…$15.00
https://doi.org/10.1145/3131542.3131556

VEHICLES AND DATA TRAFFIC

Data Traffic from Vehicles
In this paper, we focus on the data transmitted between the base stations and the devices inside vehicles. People travelling in vehicles use various services with mobile

devices. In addition, vehicles are equipped with devices (vehicular devices) to obtain data, control the vehicle and provide services. For example, sensor data can be sent to the cloud, processed, and then used for various purposes. All this requires connectivity and generates data traffic (Figure 1).

In general, this data traffic exhibits the following characteristics:

- Unbalanced, heterogeneous data types. The multimedia data for personal entertainment is generally transmitted in bursts and the amount of data is high. On the other hand, although the devices equipped in vehicles produce less data, the number of these devices can be higher and they typically generate data continuously during the whole driving period.
- Periodicity. Personal activity may exhibit regularity, based on daily routines. Therefore, the data may exhibit corresponding periodic features.
- Geographical variance. Since personal activities and routines are typically associated with locations, the data traffic varies between locations as well.

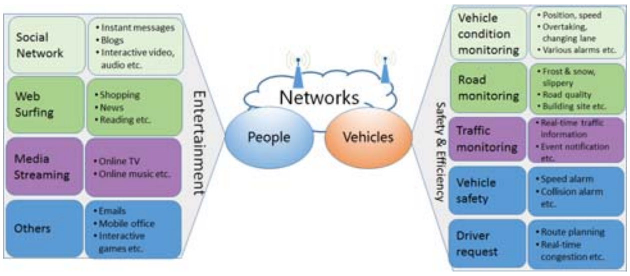


Figure 1. Various data sent from and to devices in vehicles.

To summarize, the data traffic aggregated on each base station depends on the number of vehicles on nearby roads and the data traffic from these vehicles. Moreover, the behavior and the number of people on board affect the data traffic generated by a single vehicle.

Data Traffic Modelling: State of the Art

Traffic modelling for IP-based Internet has been studied actively to determine the performance requirements that applications set to the network. Some early studies reported that IP traffic has the characteristics of self-similarity and Long-Range Dependence (LRD) [1-2]. The multifractal property of network traffic is discussed in [3]. Furthermore, analytically simple procedures, that are based on Markovian model, can be used for calculating network performance and predicting future network traffic [7][8].

Recently, due to their simple structure and tractable complexity, Hidden Markov Models (HMMs) [9] and Hidden semi-Markov Models (HsMMs) [10] have been used to model network traffic, such as source traffic [11], the arrival process of Web traffic [12], clustering and classifying workload patterns [13], and characterizing second order self-similar network traffic [14]. However, in these models the entire traffic process is assumed to be controlled by a single Markov or semi-Markov chain, which does not comply with the behavior of actual network traffic. In [15], a Nested Hidden Markov Model (NHMM) with variable state-duration is proposed to model the time-varying oscillation behavior of Internet traffic: the first layer hidden Markov chain with variable state-duration controls the time-varying oscillatory process and the second layer hidden Markov chain describes the local fluctuations within one oscillation stage. We share this idea to use two-level nested hidden Markov chain to model the periodical behavior of the data traffic. However, we use the arrival rate of the clustered data to describe the second level states and modified piecewise aggregation method to obtain the first level state, which is more related with the aggregated data traffic.

Several traffic models have been proposed for machine-to-machine (M2M) and ubiquitous sensor networks (USN). Poisson arrival process has been used to model the traffic of each individual sensor node [4], and the On/Off model for USN traffic [16]. M2M and smartphone traffic has been compared in several aspects including temporal traffic patterns, device mobility, application usage, and network performance [5]. The authors show that M2M traffic typically exhibits different diurnal patterns, and M2M more likely generates synchronized traffic in bursts. In addition, wavelet transformation has been used to decompose device model and subscriber time series. Our collected data traffic illustrates similar features as the datasets used in these works. However, we use two–level hidden Markov states to analyze the features of data traffic from different devices.

In general, the traffic model of a single application type has been considered in the context of M2M, such as telemetry [5], video surveillance [17], and YouTube [5]. In our early research [18], we proposed a model for aggregated IoT traffic. However, the aggregated traffic of M2M and H2H (human-to-human) has not been considered in earlier research from the view point of moving vehicles.

DATA COLLECTION FROM VEHICLES

Data collection from vehicles is challenging, because of the number and dynamicity of vehicles, the restrictions from the network operators, and privacy and security requirements of users. Especially, it is difficult to access the aggregated data through monitoring and statistical methods. To tackle this challenge, we reproduce the aggregated data traffic from vehicles through simulations, taking into account both the mobility of the vehicles in the real world and typical application usage patterns. This approach can be used in future scenarios as well, when there is no data to collect.

Vehicle Mobility

To acquire the mobility information of the vehicles, we use datasets recording the movements of vehicles that have been collected from the real world by other research institutions.

To have sufficient spatial and temporal span in the analysis, we selected two datasets for this research. The first one is the non-open taxi dataset from Oulu city, Finland [19], recording the movement of several hundred taxies during a period of several months in the city area of Oulu. The second one is the Cologne mobility dataset [20], recording the movement of 700 000 vehicles in an area of about 400 square kilometers for 24 hours. We use the combination of OpenStreetMap (OSM) [21], and Simulation of Urban Mobility (SUMO) [22] to process the datasets and to generate the mobility model of vehicles in the ns-3 [23] environment.

The OSM road information is generated and validated by satellite images and GPS traces, and is commonly regarded as the highest quality road data publicly available. SUMO is an open-source traffic simulator with continuous space and discrete time. SUMO is capable of importing OSM maps. We use OSM to create the real-world maps and use SUMO to generate vehicular nodes and traffic, and mobility models in the real-world road maps according to the datasets. We import the results to ns-3.

Figure 2 illustrates the positions of the base stations we generated according to the data given by [20] for the data collection. However, for the Oulu dataset, we distribute the base stations uniformly, as shown in Figure 3, as we had no knowledge of the base station locations in the City of Oulu.

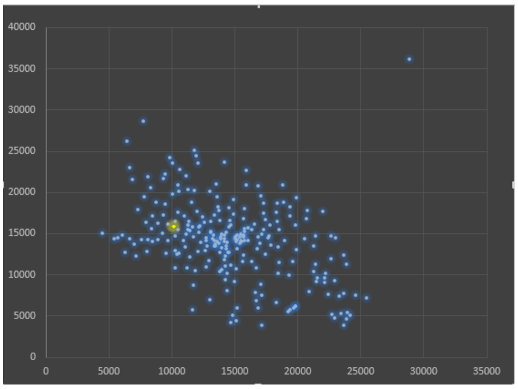


Figure 2. The distribution of base stations of Cologne

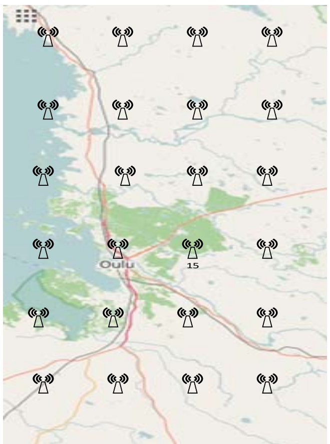


Figure 3. The distribution of base stations of Oulu

Mobile User Applications and Vehicular Devices

A traffic model might be simple to create for a single application. However, when data from multiple applications from many vehicles is mixed, together with the movement of the vehicles, the traffic characteristics becomes complicated.

In order to characterize the aggregated data traffic, we generated a large set of data in the network simulator ns-3. We simulated major applications used by mobile users and the data produced by vehicles. To produce realistic data, we generated the data according to the application characteristics reported in the literature [24-29].

Parameter	**Mean**	**Median**	**Standard deviation**	**Fitted distribution**
Main object size	31561 bytes	19471 bytes	49219 bytes	Weibull(λ=28242.8 k=0.814944)
Main object no.	2.19	1	2.63	Lognormal(μ=0.473 844,σ=0.688471)
Inline object size	23915 bytes	10284 byte	128079 bytes	Lognormal (μ=9.17979, σ= 1.24646)
Inline object no.	31.93	22	37.65	Exponential (μ =31.9291)
Reading time	39.7 s	—	324.92s	Lognormal(μ=−0.49 5204,σ=2.7731)

Table 1. Traffic pattern of Web applications [24].

We simulated three types of applications widely used by the current mobile users: web surfing, social networks, and video streaming. Table 1 shows the parameters used for simulating the web surfing application. A web page consists of one main object and several inline objects. The sizes of the main objects and inline objects are fitted by Weibull and Lognormal distributions, respectively, and the number of the main objects and inline objects follow the Lognormal and Exponential distributions, respectively. The reading time follows the Lognormal distribution with different parameters.

Parameter	Distribution & Parameter	
QQ message length L_M	Generalized extreme value	k=0.691 μ =282.413 σ =66.363
QQ message lasting time T_M	Lognormal	μ =4.487 σ = 0.4819
QQ packet length L_P	Generalized extreme value	k=1.116 μ = 121.866 σ = 41.265
Packet arrival interval T_L	Generalized Pareto	k=0.8875 σ = 20.406 θ = 0.03

Table 2. Traffic pattern of Instant Messaging QQ [25].

QQ is an Instant Messaging software widely used with mobile phones. It supports online chatting, audio and video and file transmission. Table 2 illustrates the traffic pattern of QQ used in our simulation. For the video streaming applications, we selected You Tube with progressive download with the parameters shown as Table 3, by analyzing the data in [26].

Progressive download	Time (s)	Download rate (kbyte/s) & distribution
Burst phase	5.5	609
Throttling(video)	127.5	Normal(12.91,4.28)
Throttling(audio)	127.5	Normal(12.77,2.23)
Download finished	68.5	0

Table 3. Traffic pattern of YouTube streaming.

We generated the data pattern to simulate the data produced by vehicles and their devices (Table 4) based on [27] and [28].

Message type	**Device no.**	**Packet size (byte)**	**Sending interval**
Location	3	183	2 sec.
Direction	3	86	2 sec.
Real time traffic info.	1	99	2 sec.
Intersection collision warning	1	81	6 sec.
Vehicle diagnostics and maintenance	4	83	5 sec.
Traffic signal violation warning	4	89	6 sec.
Surveillance info.	2	Gamma(12.4 6,1.18)	24 frames/s
Route planning request	1	116	Normal(20,4)
Vehicle diagnostics and maintenance	1	140	Normal1000, 200)
Pedestrian crossing info. at designated intersections	1	81	Exp (λ=4.0)
Accident report	1	124	4 hour

Table 4. Traffic pattern of vehicle related data.

In addition, user behavior varies during a day. For example, users are more likely to surf the web during the day, watch

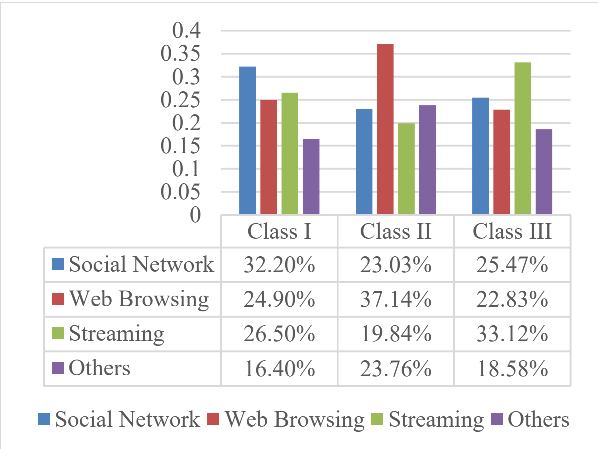


Figure 4. Ratio of applications at different times during a day

video in the night, and chat with others during the rush hour just before or after work. We set the ratio of the number of different applications during the different times of a day, as shown in Figure 4, by analyzing the data in [29].

Aggregated Data

In short, we generated vehicle traces for the ns-3 simulation from the Cologne and Oulu datasets using OSM and SUMO. Then we simulated in ns-3 various application data from each vehicle based on the traffic patterns shown in Tables 1-4. The data traffic was sent to and received from the network through the nearest base station. As a result, we obtained data traffic for each base station. Figure 5 illustrates the aggregated data of base station 15 (see Figure 3) for 7 days in the City of Oulu. Here we can see that the data traffic roughly repeats the same pattern each day.

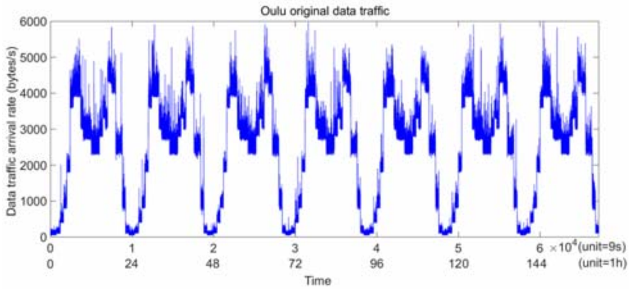


Figure 5. Aggregated data for one week in Oulu.

Figure 6 illustrates the aggregated data of the base station in Cologne (marked in Figure 2) for 24 hours. Here we can see that the two peaks of data traffic are roughly at 6.30am-7.30am and 17.00pm-19.00pm, which would probably coincide with rush hours.

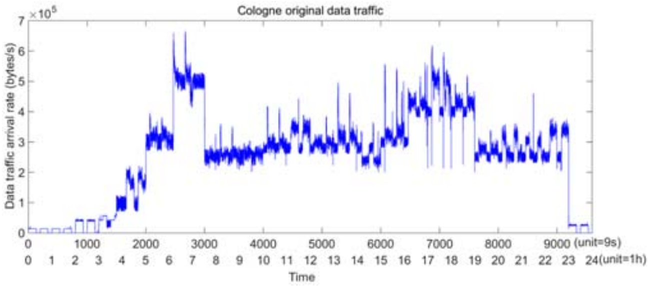


Figure 6. Aggregated data of 24 hours in Cologne

TWO-LEVEL HIDDEN MARKOV MODEL

Having the data aggregated at each base station, we analyze the characteristics of the data and use a two-level Hidden Markov Model (HMM) to model the data.

Data Characteristics

One can observe that the aggregated data shows periodicity and two-stage feature at large temporal scale. The data traffic has a crest and a valley in a period of one day. Moreover, in one day, the traffic shows stage feature.

To observe the small-scale features of the data traffic, we calculate the Probability Density Function (PDF) of the data traffic for 24 hour intervals, as shown in Figure 7. Here we can see that the data traffic shows a feature of levels, i.e., the arrival data rate at base stations concentrate on certain stages.

To confirm this feature, we drew also the PDF of the data traffic for 7 days in Oulu, as shown in Figure 8. We observe that the data traffic fluctuates at a small temporal scale and the data arrival rates concentrates on different stages.

Two-level Model

Next, we process the collected data to determine state parameters. We define two sets of states according to the fluctuation and stage features at large and small temporal scales. The first level is at the Large Scale (LS) level. After clustering the fluctuation stage using a modified Piecewise Aggregate Approximation (PAA) method [30], we get a discrete time series.

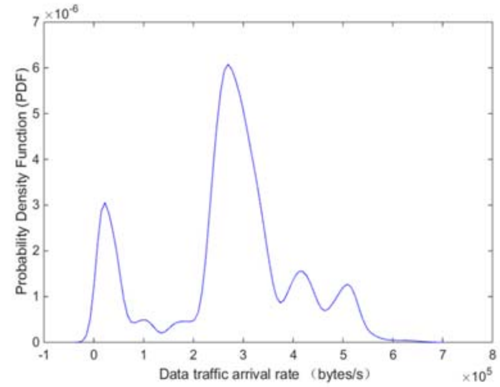


Figure 7. PDF of data traffic of in Cologne for 24 hours

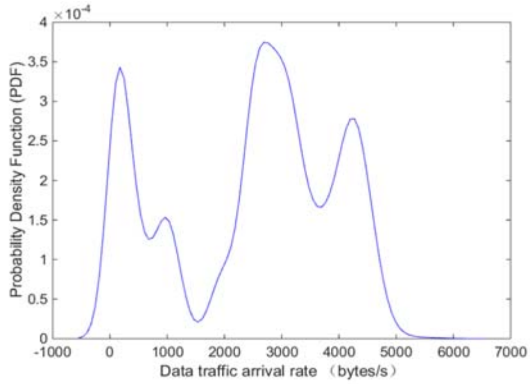


Figure 8. PDF of data traffic of in Oulu for 7 days

Let $Q^L = \{1,2,\dots,|Q^L|\}$ denote the collection of the time series at the LS level, $|Q^L|$is the number of the divided time period. q_t^L is the LS level state at time *t*. $\pi_i^L \equiv \mathrm{P}(q_1^L = i)$, $i \in Q^L$, is the probability when initial state is equal to *i*. $a_{ij}^L = P(q_t^L = j|q_{t-1}^L = i)$ is the transition probability from the state *i* to state *j*. Thus, $\{q_1^L, q_2^L, \dots, q_T^L\}$ consists of a state chain at the LS level.

The second level is at Small Scale (SS) level, it is associated with each state at the LS level. Based on the clustered LS level sequence, K-means [31] algorithm is used to cluster the traffic fluctuation in the small scale. Let $Q^{S|i} = \{1,2,\dots,|Q^{S|i}|\}$, $i \in Q^L$ be the collection of data arrival rate at the SS level when LS state is i . q_t^S is the SS level state at time *t*. $\pi_m^{S|i} \equiv P(q_t^S = m|q_t^L = i, q_{t-1}^L \neq i)$ is the probability when initial state is equal to *m* and the LS state is *i*. $a_{mn}^{S|i} = P(q_t^S = n|q_{t-1}^S = m, q_{t-1:t}^L = i)$ is the transition probability from SS state *m* to SS state *n* when the LS state is *i*. Thus, given LS state $q_{t:t+\tau}^L = i$, $\{q_t^S, q_{t+1}^S, \dots, q_{t+\tau}^S\}$ consists its internal SS state.

In addition, assume $O = \{o_1, o_2, \dots, o_T\}, o_t \in \{1,2,\dots,\}$ is the observation state sequence for a period of T, o_t is the data arrival rate at time *t*. $b_{mk}^{S|i} \equiv P(o_t = k|q_t^S = m, q_t^M = i)$ is the probability of arrival data rate is equal to *k* when LS state is *i* and SS state is *m*.

Therefore, the model can be described using $\lambda = (\pi_i^M, a_{ij}^M, \pi_m^{S|i}, a_{mn}^{S|i}, b_{mk}^{S|i})$. Here both the LS and SS chains are hidden, only the observed layer are known from the collected data.

To formulate the two level model, we assume that the LS and SS chains have the no aftereffect property of a Markov chain. We also assume the outputs are only controlled by the SS process and conditionally independent when the underlying states are given. Hence, our two level model becomes the Hierarchical Hidden Markov Model, i.e., TL-HMM.

Therefore, we can use Baum-Welch algorithms to calculate the parameters of the model $\lambda = (\pi_i^M, a_{ij}^M, \pi_m^{S|i}, a_{mn}^{S|i}, b_{mk}^{S|i})$ under the observation sequence$O = \{o_1, o_2, \dots, o_T\}$.

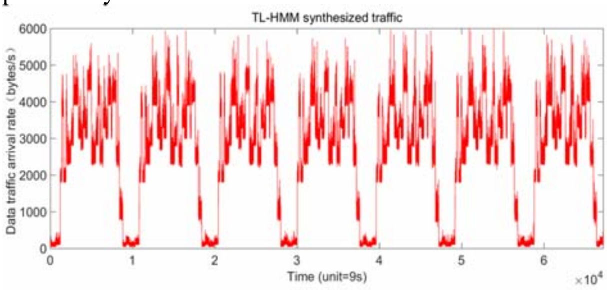
EVALUATIONS

We evaluated the proposed model by comparing the data traffic generated according to the model with the original simulation data. Figure 9 (a) and (b) illustrate the data traffic of one week in Oulu synthesized with the model. The initial probability matrix values were set randomly. Figure 9 (c) is the collected data. We also synthesize the traffic using HMM (i.e., one level), as shown in (d). Here we can see that the synthesized results varies with the initial probability matrix, but with proper values, the model can well represent the aggregated data traffic in one base station from a number of moving vehicles.

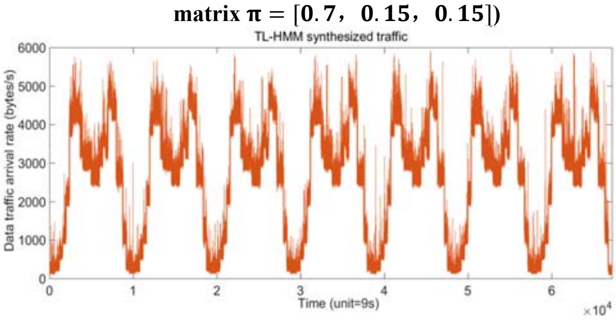
Figure 10 (a) and (b) illustrates the synthesized traffic of 24 hours in Cologne using different initial probability matrixes, and (c) is the collected data. Here we can see that through the two level model, both the large scale and small scale features can be characterized. Again, the results depends on the selection of the initial probability matrix of the TL-HMM.

We calculated also the statistical parameters, as shown in Table 5 and Table 6. Here we select the initial probability matrix of $\pi = [0.8, 0.1, 0.1]$ for Oulu data and $\pi = [0.85, 0.10, 0.05]$ for Cologne data which can better model the traffic. Figure 11 (a) and (b), and Figure 12 (a) and (b) show the PDF and CDF for the two datasets.

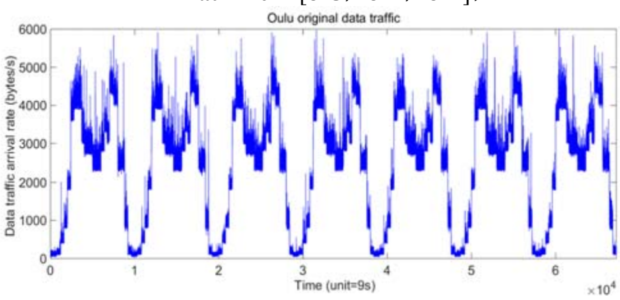
Here we can see that mean absolute percentage error of TL-HMM is less than 10%. Compared with the HMM, the proposed two-level HMM model provide a better method to model the data traffic. However, the accuracy of the model varies with the selection of the initial probability matrix. In the above tests, we selected random values. We are now investigating methods to find optimal values for the initial probability matrix.



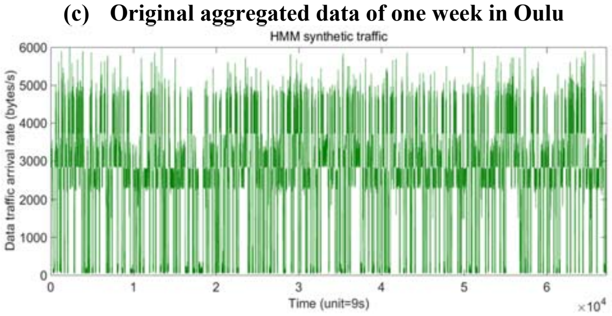
(a) Synthesized data using TL-HMM (initial probability matrix $\pi = [0.7, 0.15, 0.15]$)



(b) Synthesized data using TL-HMM (initial probability matrix $\pi = [0.8, 0.1, 0.1]$)



(c) Original aggregated data of one week in Oulu

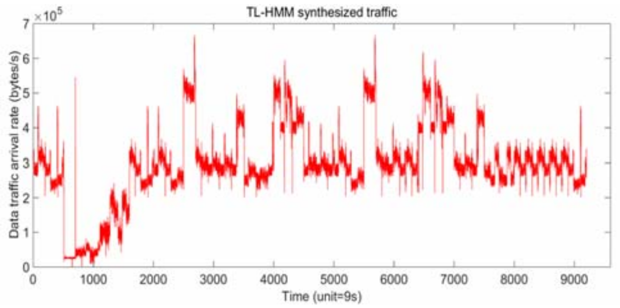


(d) Synthesized data using one level HMM

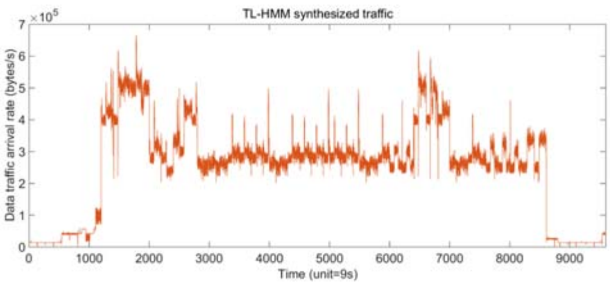
Figure 9. Comparison of original and synthesized data of one week in Oulu

Parameters / Model	Mean absolute error (Byte)	Mean absolute percentage error (%)	Standard deviation (Byte)
TL-HMM	85	8.7	231
HMM	337	15	400

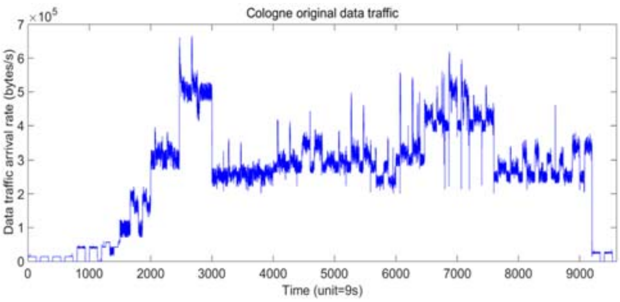
Table 5. Statistical difference of one week Oulu data



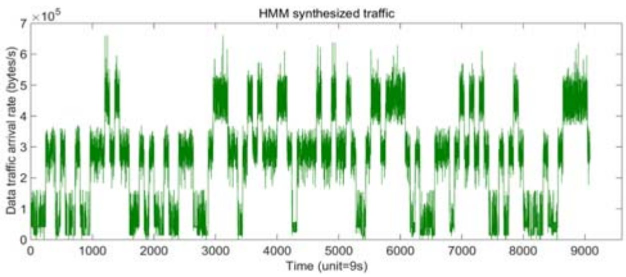
(a) Synthesized data using TL-HMM (initial probability matrix $\pi = [0.7, 0.15, 0.15]$



(b) Synthesized data using TL-HMM (initial probability matrix $\pi = [0.85, 0.10, 0.05]$)



(c) Original aggregated data of 24 hours in Cologne

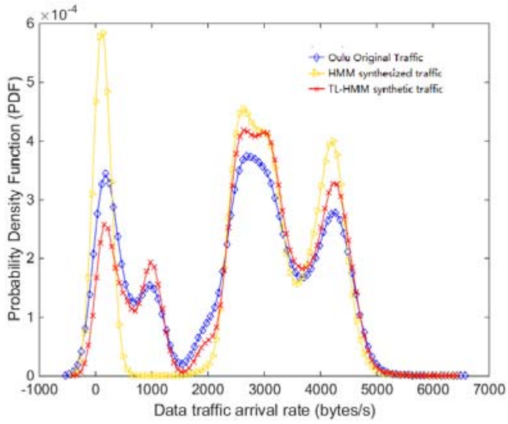


(d) Synthesized data using one level HMM

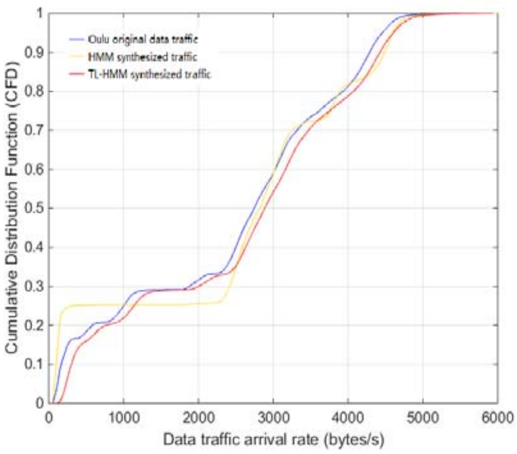
Figure 10. Comparison of original and synthesized data of 24 hours in Cologne

Parameters / Model	Mean absolute error(Byte)	Mean absolute percentage error (%)	Standard deviation (Byte)
TL-HMM	110.23	9.34	129.81
HMM	158.36	13.48	188.54

Table 6. Statistical difference of 24 hours Cologne data

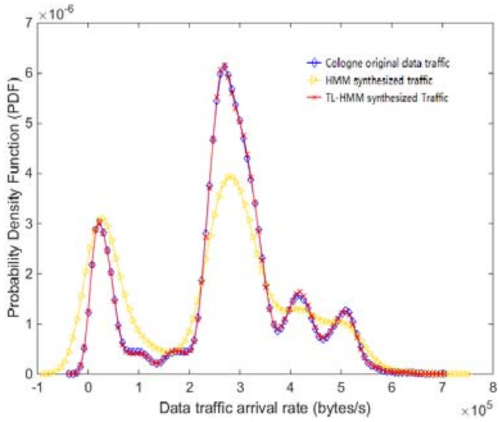


(a) Probability Density Function (PDF)

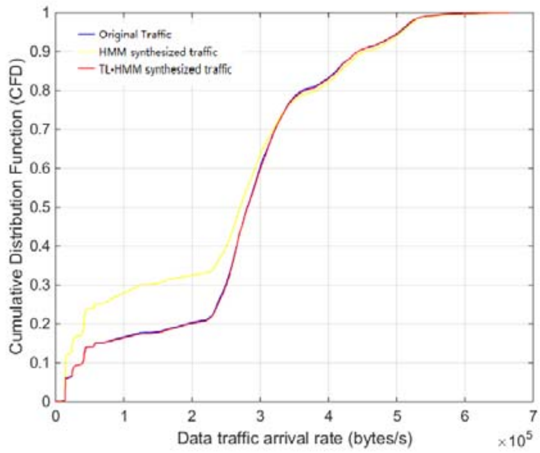


(b) Cumulative Distribution Function (CFD)

Figure 11. PDF and CDF of Oulu data



(a) Probability Density Function (PDF)



(b) Cumulative Distribution Function (CFD)

Figure 12. PDF and CDF of Cologne data

CONCLUSIONS AND FUTURE WORK

In this paper we proposed a Two-level Hidden Markov model (TL-HMM). The first level hidden Markov chain models the large scale characteristics of the data traffic, mainly influenced by the daily behavior of the people on board. The second level hidden Markov chain models the small scale characteristics of the data traffic, mainly influenced by the data produced by the vehicular devices. Our experimental results show that the proposed model can be used to characterize the data traffic from vehicles.

The model can help to estimate the network performance and predict the future network traffic considering the impacts from increasing number of vehicles. Our future work is to refine the proposed model by analyzing more datasets and studying initial probability matrix values. We will also consider how the model could be verified with measurements from a real network, for example, the 5G Test Network (http://5gtn.fi/). Moreover, we will investigate the influence of the data traffic from the vehicles on the core network and consider future scenarios for both end-user application usage and vehicular devices. Finally, combining this model with other design tools would enable more detailed analysis of the requirements that vehicles set to networks.

ACKNOWLEDGMENTS

The work is partly supported by Huawei HIRP program.

REFERENCES

1. V. Paxson and S. Floyd, “Wide-area traffic, “The failure of Poisson modeling,” IEEE/ACM Transactions on Networking, vol. 3, no. 3, pp.226–244, June 1995.
2. W. Leland, M. Taqqu, W. Willinger, and D. Wilson, “On the self-similar nature of Ethernet traffic (extended version),” IEEE/ACM Transactions on Networking, vol. 2, no. 1, pp. 1–15, Feb. 1994.
3. R. Riedi, M. Crouse, V. Ribeiro, and R. Baraniuk, “A multifractal wavelet model with application to network

traffic," IEEE Transactions on Information Theory, vol. 45, no. 4, pp. 992–1018, April 1999.

4. Obada Al-Khatib, Wibowo Hardjawana, Branka Vucetic, "Traffic Modeling for Machine-to-Machine (M2M) Last Mile Wireless Access Networks," Communications QoS, Reliability and Modelling Symposium 2014.
5. M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang and J. Wang, "Large Scale Measurement and Characterization of Cellular Machine-to-Machine Traffic," IEEE/ACM Transactions on Networking, 2013, 21(6):1960-1973.
6. G. Horváth and 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.
7. A. Andersen and B. Nielsen, "A Markovian approach for modeling packet traffic with long-range dependence," IEEE Journal on Selected Areas in Communications, vol.16, no.5, pp.719–732, June 1998.
8. L. Muscariello, M. Meillia, M.Meo and M.A. Marsan, "An MMPP-Based Hierarchical Model of Internet Traffic," in Proc. of IEEE Internationcal Conference on Communications. pp.2143-2147, 2004.
9. L. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proc. IEEE, vol. 77, no. 2, pp. 257-286, Feb. 1989.
10. S. Yu, "Hidden Semi-Markov Models," Artificial Intelligence, vol. 174, no. 2, pp. 215-243, 2010.
11. J. Doma_nska, A. Doma_nski, and T. Czacho´ rski, "Internet Traffic Source Based on Hidden Markov Model," Proc. 11th Int'l Conf. and Fourth Int'l Conf. Smart Spaces and Next Generation Wired/Wireless Networking, pp. 395-404, 2011.
12. S. Yu and H. Kobayashi, "An Efficient Forward-Backward Algorithm for an Explicit-Duration Hidden Markov Model," IEEE Signal Processing Letters, vol. 10, no. 1, pp. 11-14, Jan. 2003.
13. N. Li and S. Yu, "Periodic Hidden Markov Model-Based Workload Clustering and Characterization," in Proc. of IEEE Eighth Int'l Conf. Computer and Information Technology, pp. 378-383, 2008.
14. S. Yu, Z. Liu, M. Squillante, C. Xia, and L. Zhang, "A Hidden Semi-Markov Model for Web Workload Self-Similarity," in Proc. IEEE 21st Int'l Performance, Computing, and Comm. Conf., pp. 65-72, 2002.
15. Y. Xie, J. Hu, Y. Xiang, et al., "Modeling Oscillation Behavior of Network Traffic by Nested Hidden Markov Model with Variable State-Duration," IEEE Transactions on Parallel & Distributed Systems, 2013, 24(9):1807-1817.
16. 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, October 27-31, 2008, pp.96-100.
17. I. Petiz, P. Salvador and A. Nogueira, "Characterization and Modeling of M2M Video Surveillance Traffic," in Proc. of the Fourth International Conference on Advances in Future Internet (AFIN 2012), pp.65-70.
18. Y. Li, Y. Huang, X. Su, J. Riekki, et. al, "A Gamma-modulated Wavelet Model for Internet of Things Traffic," in Proc. of The 2017 IEEE International Conference on Communications (ICC 2017), Paris, France, May 21-25, 2017.
19. Project deliverable of Data to Intelligence. (http://www.datatointelligence.fi/)
20. Vehicular mobility trace of the city of Cologne, Germany, http://kolntrace.project.citi-lab.fr/
21. Open Street Map. http://www.openstreetmap.org
22. Simulation of Urban Mobility. http://sumo.sourceforge.net
23. ns-3: A Discrete-event Network Simulator. http://www.nsnam.org/
24. Rastin Pries, Zsolt Magyari and Phuoc Tran-Gia, "An HTTP Web Traffic Model Based on the Top One Million Visited Web Pages," Next Generation Internet, 2012:133-139.
25. H. Lu, C. Yu and X. Gou, "Analysis of Traffic Model and Self-similarity for QQ in 3G Mobile Networks," in Proc. of The International Conference on Advanced Intelligence & Awareness Internet, 2011:131-135.
26. Géza Horváth and Péter Fazekas, "Modelling of YouTube Traffic in High Speed Mobile Networks," in Proc. of European Wireless Conference, 2015.
27. S. Olariu, M. C. Weigle, (Ed.), "Vehicular Networks, From Theory to Practice," CRC Press.
28. J. Zhao, "Analysis of the Characteristics of M2M Traffic," Master thesis, Southwest Jiaotong University, 2014.
29. Sandvine Intelligent Broadband Networks, Global Internet Phenomena Report 1H 2014.
30. E. J. Keogh, M.J. Pazzani, "Scaling up dynamic time warping for datamining applications," In Proc. of Internal Conference on Knowledge Discovery & Data Mining. ACM, 2000:285-289.
31. C. Ding and X. He, "K-means Clustering via Principal Component Analysis," in Proc. of Int'l Conf. Machine Learning (ICML 2004), July, 2004, pp. 225–232.