www.dergipark.gov.tr
ISSN:2148-3736
El-Cezerî Fen ve Mühendislik Dergisi
Cilt: 7, No: 1, 2020 (295-303)

El-Cezerî Journal of Science and Engineering
Vol: 7, No: 1, 2020 (295-303)

DOI : 10.31202/ecjse.640824

ECJSE

Research Paper / Makale

Data-Driven Modelling and Prediction of CoAP Throughput in a Grid Network Topology

Alper Kamil DEMIR, Fatih ABUT

Adana Alparslan Türkeş Science and Technology University, Faculty of Engineering, Computer Engineering Department, Adana, Turkey
akdemir@atu.edu.tr

Received/Geliş: 31.10.2019 **Accepted/Kabul:** 31.12.2019

Abstract: In this study, we propose new models for predicting the average throughput in a 4x4 grid Constrained Application Protocol (CoAP)-based IoT network using Support Vector Machine (SVM) and Multiple Linear Regression (MLR). Two different CoAP congestion control mechanisms have been considered: the default CoAP congestion control (CC) and the CoAP Simple Congestion Control/Advanced (CoCoA). On the client-side, we run 3, 6, 9, 12 or 15 CoAP clients requesting packets, sized with 12, 24, 36 or 48 bytes, from different CoAP servers over 4x4 grid IoT network configured with packet delivery ratios of 90, 95 or 100. In total, 60 different experimental scenarios, each of which was run 10 times to determine the average throughput of default CoAP CC and CoCoA clients, were created. Using 10-fold cross-validation, the performance of the prediction models has been evaluated using several performance metrics. The results show that combining packet delivery ratio and number of concurrently sending clients in a model leads to the highest correlation with the average CoAP throughput of the IoT network. Particularly, this model produces the lowest prediction error among all SVM-based and MLR-based models, regardless of whether the default CoAP CC or CoCoA is used as the congestion control mechanism.

Key words: Internet of Things, Throughput Prediction, CoAP, CoCoA, Contiki OS

Bir Izgara Ağı Topolojisi Üzerinde CoAP'ın Veriye Dayalı Modellemesi ve Verimliliğinin Tahmini

Özet: Bu çalışmada, Destek Vektör Makinesi (SVM) ve Çoklu Doğrusal Regresyon (MLR) kullanılarak 4x4'lük ızgara topolojisi üzerinde Kısıtlı Uygulama Protokolü (CoAP) tabanlı bir IoT ağındaki ortalama verimi tahmin etmek için yeni modeller önerilmektedir. İki farklı CoAP tıkanıklık kontrol mekanizması dikkate alınmıştır: mevcut CoAP tıkanıklık kontrolü (default CoAP CC) ve CoAP Simple Congestion Control/Advanced (CoCoA). İstemci tarafında, 90, 95 veya 100 paket teslimat oranları ile yapılandırılmış 4x4'lük ızgara IoT ağı üzerindeki farklı CoAP sunucularından, 12, 24, 36 veya 48 bayt boyutunda paket talep eden 3, 6, 9, 12 veya 15 CoAP istemcisi çalıştırılmıştır. Toplamda, mevcut CoAP CC ve CoCoA istemcilerinin ortalama verimini belirlemek için her biri 10 kez çalıştırılan 60 farklı deneysel senaryo oluşturulmuştur. 10 katlı çapraz doğrulama kullanılarak, tahmin modellerinin performansı çeşitli performans ölçümleri kullanılarak değerlendirilmiştir. Sonuçlar, paket teslim oranının ve aynı anda gönderen istemci sayısının aynı modelde birleştirilmesinin IoT ağının ortalama CoAP verimi ile en yüksek korelasyona sahip olduğunu göstermektedir. Özellikle, bu model varsayılan CoAP CC veya CoCoA'nın tıkanıklık kontrol mekanizmalarından bağımsız olarak, tüm SVM tabanlı ve MLR tabanlı modeller arasında en düşük tahmin hatasını üretmektedir.

Anahtar kelimeler: Nesnelerin İnterneti, Verimlilik Tahmini, CoAP, CoCoA, Contiki OS

How to cite this article
Demir, A. K., Abut, F., "*Data-Driven Modelling and Prediction of CoAP Throughput in a Grid Network Topology",* El-Cezerî Journal of Science and Engineering, 2020, 7(1); 295-303.

Bu makaleye atıf yapmak için
Demir, A. K., Abut, F., "*Bir Izgara Ağı Topolojisinde CoAP'ın Veriye Dayalı Modellemesi ve Verimliliğinin Tahmini",* El-Cezerî Fen ve Mühendislik Dergisi 2020, 7(1); 295-303.

1. Introduction

The Internet of Things (IoT) heralds the vision of future Internet where increasing number of physical devices, embedded with sensors, actuators, communication unit, electronics, software and protocols, exchange information about themselves and their surroundings within existing Internet infra- structure [1]. The IoT transforms these physical devices into smart things where a new generation of applications in city, transportation, industry, factory, market, school, vehicles, home, agriculture, healthcare, grid, power plant, aerospace, aviation and other domains will enrich our life [2, 3].

As IoT devices are equipped with constrained energy, processing unit, memory and communication capacity, a new generation of software and networking protocols are being developed and standardized. In this sense, the Internet Engineering Task Force (IETF) constructed the Constrained Application Protocol (CoAP) as an application layer protocol for constrained IoT devices [4]. As the Hypertext Transfer Protocol (HTTP) of web on the Internet become ubiquitous in most applications, the CoAP was also developed as web transfer protocol to satisfy the needs of constrained IoT devices. Thus, the CoAP has very similar features to HTTP.

The CoAP diverges from the HTTP as constrained IoT devices require low overhead protocols. Thus, unlike HTTP-based protocols based on complex TCP, CoAP operates over simple UDP which does not provide any congestion control (CC) service. Hence, the CoAP employs congestion control services. However, instead of using complex congestion control as in TCP, the core CoAP specification leverages a simple CC mechanism utilizing retransmission timeout (RTO) with binary exponential backoff (BEB) [5]. As lost packets are retransmitted at exponentially growing time, default CoAP CC is very basic and insensitive to network dynamics. So, default CoAP CC can be categorized as conservative. For not adjusting its protocol operation to dynamic network conditions, default CoAP CC may consequently underperform. Thus, core CoAP specification is receptive to novel CC mechanisms leveraging dynamic network conditions. Accordingly, CoAP Simple Congestion Control/Advanced (CoCoA) [6] is proposed to improve the default CoAP CC. CoCoA harnesses round-trip time (RTT) measurements, adaptive RTO back off computations, and RTO aging methods to further improve the performance of default CoAP CC. More detailed information about the default CoAP CC and CoCoA can be found in [7].

As IoT devices are equipped with constrained communication capacity, the bandwidth of an IoT network is low in general. This situation requires careful allocation of bandwidth for IoT applications. Particularly, bandwidth and Quality of Service (QoS) sensitive IoT applications rely on accurate prediction of IoT network throughput. In this context, IoT network throughput is defined as the number of messages successfully delivered per unit time. However, it is not straightforward to predict the throughput of the default CoAP CC and CoCoA as such congestion control mechanisms are based on adaptive RTO measurements. As a result, the CoAP CC and CoCoA mechanisms result with throughput variations depending on RTO measurements.

In this paper, we extend our previous study in [8] and predict the throughput of CoAP clients using two different congestion control mechanisms. More specifically, in [8], seven different models were developed to predict the throughput of default CoAP CC using SVM and MLR. To create the ground-truth dataset, totally 60 different experimental scenarios were conducted to measure the default CoAP CC throughput which acted as reference values for evaluating the accuracy of the predictions. In this study, in addition to the default CoAP CC measurements presented in [8], the same 60 experimental scenarios were conducted again to collect the ground-truth measurements of CoCoA throughput. Consequently, in this study, also the CoCoA mechanism has been considered in the development and evaluation of prediction models. The contributions of the paper can be summarized as follows:

- For the first time in literature, we propose new models for predicting the throughput of CoAP by machine learning based approach in IoT networks. A ground-truth dataset containing 60 samples related to various experiment scenarios was created by varying number of clients, packet size and packet delivery ratio (PDR) of the IoT network. The throughput of each CoAP client was measured using the default CoAP CC and CoCoA.
- By using the created dataset, 14 models for predicting the throughput of default CoAP CC and CoCoA were developed based on SVM and MLR.
- The performance of the prediction models was evaluated by using 10-fold cross-validation and various evaluation metrics.

The rest of the paper is organized as follows. Section 2 gives an overview of related works. Section 3 gives the details of testbed setup and dataset generation. Section 4 presents the evaluation methodology. Results and discussion are given in Section 5. Finally, in Section 6 the paper is concluded along with future directions.

2. Related Works

Throughput prediction has been an active research in cellular/mobile networks [9, 10], data center networks [11, 12], wireless networks [13, 14] and wide area networks [15, 16]. In the context of cellular/mobile networks, seven TCP throughput prediction algorithms are examined in [9]. These algorithms are based on either throughput, multiple linear regression, neural network regression, or support vector regression. On the other hand, in [10], instantaneous (i.e. history-less) prediction of reachable throughput of a connection over a time slice of a few seconds and based on several metrics and parameters is studied for cellular/mobile networks. Metrics and parameters that enable relatively accurate predictions (i.e. Received Signal Strength Indicator, GPS coordinates and speed) are identified to predict the throughput.

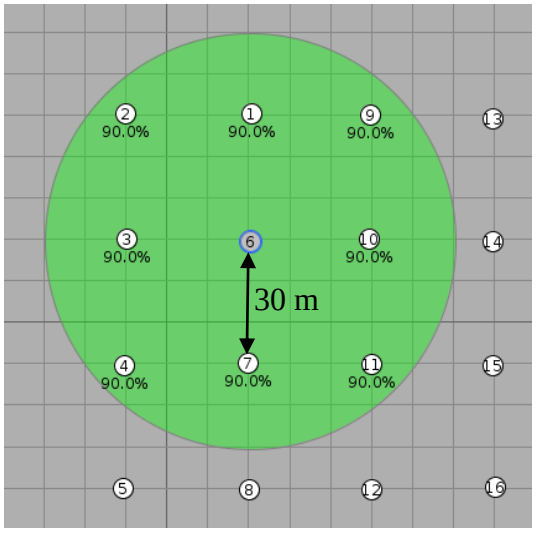
Moreover, throughput prediction in data center networks is also of research interest [11, 12]. Both studies use model-based techniques to predict the throughput. The work in [11] uses Descartes Meta-Model (DMM) whereas the work in [12] is based on Fast Factor Learning (FFL) model. Time series analysis and machine learning techniques are used in opportunistic wireless networks where the two communication endpoints cooperate in the prediction procedure, and there does not exist any network topology and traffic load knowledge [13]. In [14], the aggregated throughput of IEEE 802.11 based wireless networks is predicted using directional antenna. Particularly, a model for estimating the aggregated throughput of IEEE 802.11 based wireless network using the improved attacking case metric is derived.

DualPats [15] reveals that there is a strong correlation between TCP throughput and flow size, and statistical stability of Internet path characteristics. Based on these observations, TCP throughput of large transfers is accurately predicted by using active probing on the wide area network. A new light way method for TCP throughput prediction on wide area network environment is presented in [16]. SVM is leveraged on both prior file transfer history and measurements of simple path properties. However, despite these studies, there is no any research on predicting the throughput estimation in IoT networks. Therefore, we initiate such research by machine learning based approach for CoAP.

3. Experimental Setup and Dataset Generation

To create the dataset, we run varying number of CoAP clients residing at the Internet and requesting packet from CoAP servers residing at the 4x4 grid IoT network. The CoAP clients are programmed with Californium implementation of CoAP [17]. The numbers of clients are always equal to the

numbers of different servers residing at a 4x4 grid IoT network. Each CoAP server is programmed with Erbium implementation [18] of CoAP in Contiki OS [19]. Moreover, Erbium CoAP servers are based on UDP, uIPv6, RPL, SICSlowpan, Carrier Sense Multiple Access (CSMA) and nullRDC network stack running over IEEE 802.15.4 physical layer. The 4x4 grid IoT network is emulated at the Cooja simulator of Contiki OS. Figure 1 illustrates the emulated 4x4 grid network topology.



Overview of simulation parameters:

Packet sizes: 12, 24, 36 and 48 bytes

PDRs: 90%, 95% and 100%

Number of clients: 3, 6, 9, 12 and 15

Congestion control: Default CoAP CC and CoCoA

MAC layer: Carrier Sense Multiple Access (CSMA) and nullRDC

Experiments: In total, 60 different experiment scenarios each of which has been repeated 10 times and has a duration of 3 minutes.

Figure 1. Emulated 4x4 grid network topology and simulation parameters

The number of CoAP clients is varied as 3, 6, 9, 12 and 15. The requesting packet size of CoAP clients is varied as 12, 24, 36 and 48 bytes. Moreover, the PDR of the 4x4 grid network is varied as 90%, 95% and 100%. Consequently, totally 60 different experimental scenarios are obtained for CoAP clients. The emulations of the 60 different experimental scenarios have been repeated twice in order the measure the CoAP throughput using (a) the default CoAP CC and (b) the CoCoA mechanisms. Particularly, for each experimental scenario, we coded a shell script to run CoAP clients and created a Cooja simulation file with related packet size and PDR. All CoAP clients are run 3 minutes in parallel. Each experimental scenario is run 10 times. We used 30 PCs to run our experiments. As we have 60 experimental scenarios that run 10 times, each PC runs two experimental scenarios at different times. The throughput of CoAP clients is recorded in a text file for each experimental scenario. We also wrote a java program to retrieve the average default CC and CoCoA throughput of CoAP clients recorded in a text file. As a result, 120 different average CoAP throughput values, i.e. 60 average throughput values for default CoAP CC (1.04 ± 0.57) and 60 average throughput values for CoCoA (1.05 ± 0.60) of varying number of CoAP clients, packet size and PDR form our dataset.

4. Evaluation Methodology

For predicting the default CoAP CC and CoCoA throughput, various prediction models have been developed by using the single, double and triple combinations of the three predictor variables. Table 1 presents the prediction models along with the predictor variables that each model includes. For every model, the default CoAP CC and CoCoA throughput has been predicted using SVM and MLR separately. The values of cost (C), epsilon (ϵ) for the ϵ-insensitive loss function, and the type of kernel function are the main parameters influencing the performance of SVM-based models. The radial basis function has been utilized as the kernel function, which requires the optimization of the function parameter gamma (γ). The optimal values of the three model parameters C, ϵ, and γ have been determined using the grid search technique.

The performance of the prediction models is evaluated by calculating their root mean square errors

(*RMSE*s), mean absolute errors (*MAE*s), and mean absolute percentage errors (*MAPE*s), the formulas of which are shown in Eqs. (1) through (3), respectively.

Table 1. Overview of default CoAP CC and CoCoA throughput prediction models

Models	**Predictor Variables**
Model 1 and Model 8	Packet size
Model 2 and Model 9	PDR
Model 3 and Model 10	Number of concurrently sending clients
Model 4 and Model 11	Packet size, PDR
Model 5 and Model 12	Packet size, number of concurrently sending clients
Model 6 and Model 13	PDR, number of concurrently sending clients
Model 7 and Model 14	Packet size, PDR, number of concurrently sending clients

$$RMSE = \sqrt{\frac{\sum \left(Y - Y^{'}\right)^{2}}{n}} \tag{1}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y - Y^{'}\right|, \tag{2}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{\left|Y - Y^{'}\right|}{Y}. \tag{3}$$

In Eqs. (1) through (3), Y is the measured throughput value, $Y^{'}$ is the predicted throughput value and n is the number of samples in a test subset. The evaluation of the generalization errors of the models, on the other hand, has been conducted using 10-fold cross-validation. Consequently, for each fold, the training data included 54 samples, while the test data included 6 samples.

5. Results and Discussion

Table 1 and Table 2 show the values of *RMSE*, *MAE* and *MAPE* for all developed SVM-based and MLR-based prediction models, respectively. Based on the results obtained, the following discussions can be made for prediction of CoAP throughput:

- When all models are examined, Model 6 and Model 13 including the predictor variables PDR and number of concurrently sending clients, in general, has the lowest *RMSE*s for prediction of CoAP CC and CoCoA throughput, respectively, independent of whether they are built with SVM or MLR. Figure 2 shows the percentage decrease rates in default CoAP CC throughput of Model 6 compared to Model 1 through Model 5 using SVM and MLR. Similarly, Figure 3 shows the percentage decrease rates in CoCoA throughput of Model 13 compared to Model 8 through Model 12 using SVM and MLR. As Model 6 and Model 7 show similar performance results on prediction of default CoAP CC, only Model 6 is considered in Figure 2. Similarly, since the results of Model 13 and Model 14 are comparable, only Model 13 is considered in Figure 3. Furthermore, Figure 4 and Figure 5 show the scatter plots of actual vs. predicted default CoAP CC and CoCoA throughput values using SVM for Model 6 and Model 13, respectively.
- In contrast, Model 1 including only a single predictor variable (i.e. packet size) has the highest *RMSE*s for prediction of default CoAP CC and CoCoA throughput, independent of whether it is built with SVM or MLR.
- In general, default CoAP CC throughput prediction models yield lower *RMSE*s than CoCoA throughput prediction models, regardless of which regression method was utilized. More

specifically, as compared to *RMSE*s of CoCoA throughput prediction models, the average percentage decrement rates in *RMSE*s for default CoAP CC throughput prediction models are 23.42% and 13.98% for SVM and MLR, respectively.

- When the effect of the three predictor variables is investigated it is seen that inclusion of packet size in SVM-based and MLR-based prediction models has a negligible effect on prediction accuracy. The PDR and the number of concurrently sending clients, in contrast, have a strong effect on the prediction of default CoAP CC and CoCoA throughput. Particularly, the outcomes show that the individual addition of these variables into the prediction models comparatively leads to significantly lower *RMSE*s.
- In general, the SVM-based prediction models outperform or at least exhibit similar performance than the MLR-based prediction models for prediction of default CoAP CC and CoCoA throughput. Particularly, compared with the *RMSE*s of models developed by MLR, the percentage decrement rates in *RMSE*s of models built by SVM range from 1.72% to 53.84% in case of default CoAP CC, and from 1.67% to 27.27% in case of CoCoA.

Table 2. Results for prediction of default CoAP CC throughput using SVM and MLR

Models	**SVM**			**MLR**		
	RMSE	***MAE***	***MAPE* (%)**	***RMSE***	***MAE***	***MAPE* (%)**
Model 1	0.57	0.44	48.29	0.58	0.45	48.77
Model 2	0.55	0.42	44.60	0.54	0.44	47.11
Model 3	0.28	0.20	17.58	0.32	0.24	23.85
Model 4	0.55	0.43	45.51	0.55	0.44	47.26
Model 5	0.29	0.20	18.16	0.32	0.25	24.32
Model 6	0.12	0.07	5.80	0.26	0.20	21.95
Model 7	0.12	0.08	8.18	0.26	0.21	22.91

Table 3. Results for prediction of CoCoA throughput using SVM and MLR

Models	**SVM**			**MLR**		
	RMSE	***MAE***	***MAPE* (%)**	***RMSE***	***MAE***	***MAPE* (%)**
Model 8	0.60	0.57	53.57	0.62	0.49	54.27
Model 9	0.58	0.46	51.96	0.58	0.47	51.39
Model 10	0.34	0.20	21.54	0.38	0.25	29.04
Model 11	0.59	0.46	51.43	0.60	0.48	53.28
Model 12	0.39	0.22	23.72	0.39	0.26	29.53
Model 13	0.24	0.15	15.02	0.33	0.23	27.54
Model 14	0.25	0.14	16.15	0.33	0.24	27.89

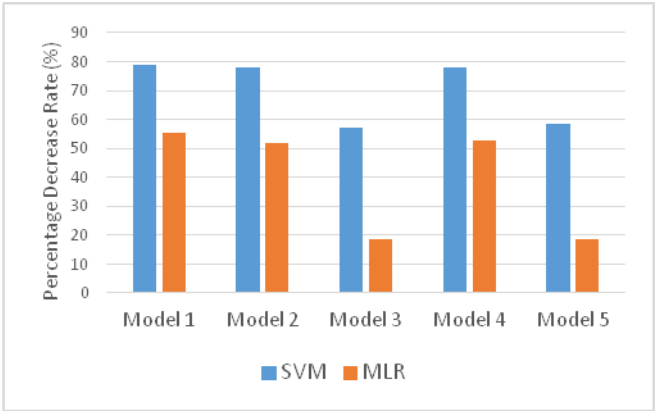


Figure 2. Percentage decrease rates in default CoAP CC throughput of Model 6 compared to Model 1 through Model 5 using SVM and MLR

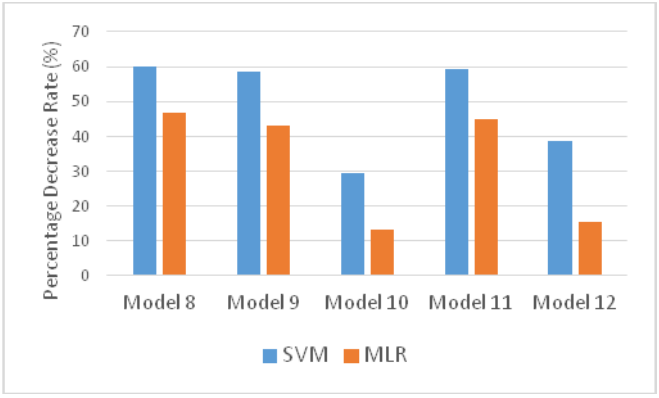


Figure 3. Percentage decrease rates in CoCoA throughput of Model 13 compared to Model 8 through Model 12 using SVM and MLR

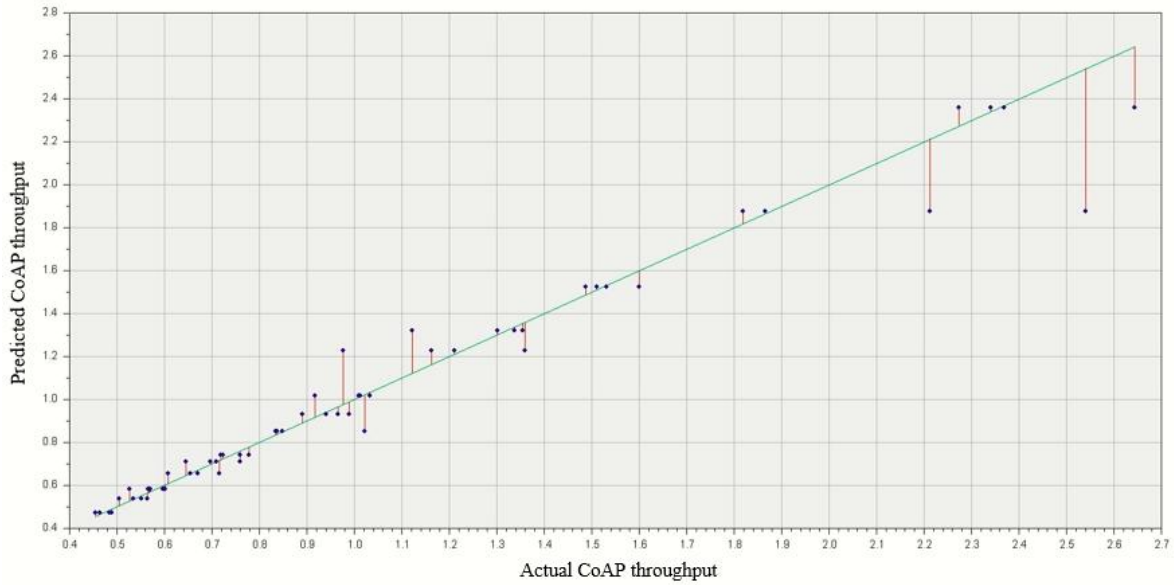


Figure 4. Scatter plot of actual throughput vs. predicted CoAP throughput for Model 6 using SVM

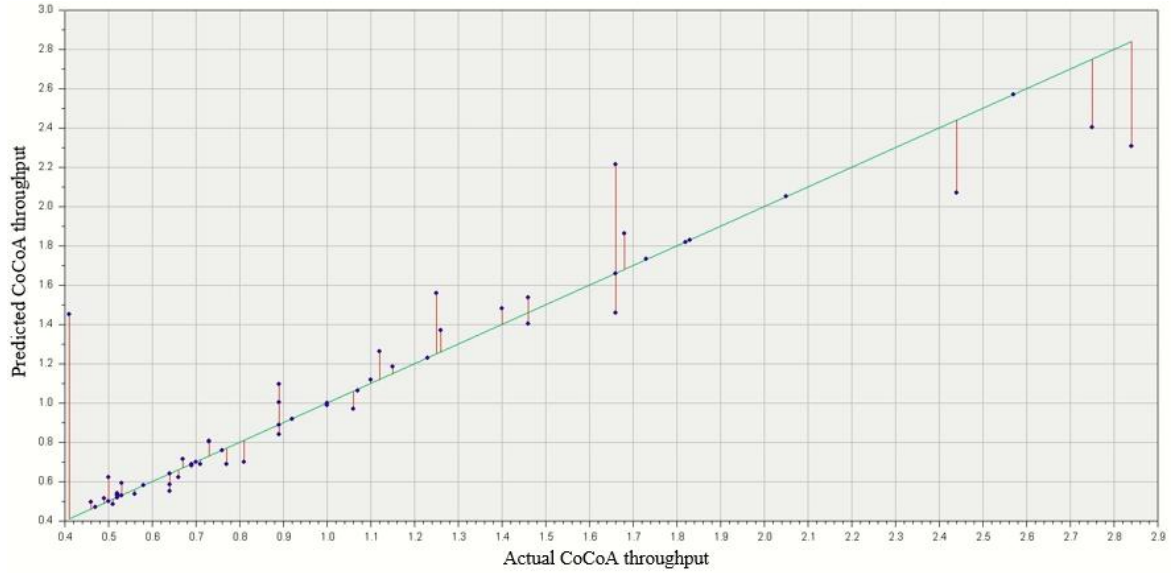


Figure 5. Scatter plot of actual throughput vs. predicted CoCoA throughput for Model 13 using SVM

6. Conclusion and Future Work

In this paper, seven new models for predicting the average throughput in a 4x4 grid CoAP-based IoT network have been developed. The created dataset used to form the prediction models includes three predictor variables, namely packet size, PDR and number of concurrently sending clients; whereas the target variables are the throughput values of CoAP clients using the default CoAP CC or CoCoA. SVM and MLR have been applied on the prediction models, and by using 10-fold cross-validation, the performance of the models has been assessed by computing several performance evaluation metrics.

The results show that combining the packet delivery ratio and number of concurrently sending clients in a model leads to the highest correlation with the average CoAP throughput of the 4x4 grid IoT network, independent of whether the default CoAP CC or CoCoA is used as the congestion control mechanism. In contrast, the packet size has the lowest relevance and correlation with the CoAP throughput. Finally, the SVM-based default CoAP CC and CoCoA throughput prediction models exhibit better or at least similar performance than the MLR-based prediction models.

This is an initial case study to show that the average CoAP throughput in a 4x4 grid topology can, in advance, be predicted with acceptable error rates using SVM and MLR. Several research directions for future work are available for this prediction field. Further scenarios with more complex grid topologies can be considered to extend the training dataset and to gain the capability of predicting the CoAP throughput for various-sized grid network topologies. Other candidate potential predictors of CoAP throughput such as the average round-trip time, jitter or number of hops can be investigated. Other promising regression methods such as deep learning and artificial neural networks can be utilized to investigate whether more accurate prediction models can be developed.

References

[1] Atzori, L., A. Iera, G. Morabito, The Internet of Things: A survey, Computer Networks, 54(15), 2787–2805, Oct. 2010.

[2] Bandyopadhyay, D., J. Sen, Internet of Things: Applications and Challenges in Technology and Standardization, Wireless Personal Communications, 58(1), 49–69, May 2011.

[3] Li, S., L. Da Xu, S. Zhao, The Internet of Things: A Survey, Information Systems Frontiers, 17(2), 243–259, Apr. 2015.

[4] Shelby, Z., K. Hartke, C. Bormann, Constrained Application Protocol. RFC 7252, 2014. [Online]. Available: https://tools.ietf.org/html/rfc7252.

[5] Betzler, A., C. Gomez, I. Demirkol, J. Paradells, CoAP congestion control for the internet of things, IEEE Communications Magazine, 54(7), 154–160, Jul. 2016.

[6] Bormann, C., A. Betzler, C. Gomez, I. Demirkol, CoAP Simple Congestion Control/Advanced, 2018. [Online]. Available: https://tools.ietf.org/html/draft-ietf-core-cocoa-03.

[7] Betzler, A., C. Gomez, I. Demirkol, J. Paradells, CoCoA+: An advanced congestion control mechanism for CoAP, Ad Hoc Networks, 33, 126–139, Oct. 2015.

[8] Demir, A. K., F. Abut, Data-Driven Modelling and Prediction of CoAP Throughput in A Grid Network Topology, in Cilicia International Symposium on Engineering and Technology, 2018, 69–73.

[9] Liu, Y., J. Y. B. Lee, An Empirical Study of Throughput Prediction in Mobile Data Networks, in IEEE Global Communications Conference (GLOBECOM), 2015, 1–6.

[10] Samba, A., Y. Busnel, A. Blanc, P. Dooze, G. Simon, Instantaneous throughput prediction in cellular networks: Which information is needed?, in IFIP/IEEE Symposium on Integrated Network and Service Management (IM), 2017, 624–627.
[11] Rygielski, P., S. Kounev, S. Zschaler, Model-based throughput prediction in data center networks, in IEEE International Workshop on Measurements & Networking (M&N), 2013, 167–172.
[12] Gao, K., J. Zhang, Y. R. Yang, J. Bi, Prophet: Fast, Accurate Throughput Prediction with Reactive Flows, in International Conference on Computer Communications (INFOCOM18), 2018, 1–9.
[13] Mirza, M., K. Springborn, S. Banerjee, P. Barford, M. Blodgett, X. Zhu, On The Accuracy of TCP Throughput Prediction for Opportunistic Wireless Networks, in 6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, 2009, 1–9.
[14] Kandasamy, S., R. Morla, P. Ramos, M. Ricardo, Predicting throughput in IEEE 802.11 based wireless networks using directional antenna, Wireless Networks, 1–18, Nov. 2017.
[15] Lu, D., Y. Qiao, P. A. Dinda, F. E. Bustamante, Characterizing and Predicting TCP Throughput on the Wide Area Network, 25th IEEE International Conference on Distributed Computing Systems (ICDCS'05), 414–424, 2005.
[16] Mirza, M., J. Sommers, P. Barford, X. Zhu, A machine learning approach to TCP throughput prediction, IEEE/ACM Transactions on Networking, 18(4), 1026–1039, 2010.
[17] Kovatsch, M., M. Lanter, Z. Shelby, Californium: Scalable cloud services for the Internet of Things with CoAP, in International Conference on the Internet of Things (IoT), 2014, 1–6.
[18] Kovatsch, M., S. Duquennoy, A. Dunkels, A Low-Power CoAP for Contiki, in IEEE Eighth International Conference on Mobile Ad-Hoc and Sensor Systems, 2011, 855–860.
[19] Dunkels, A., B. Gronvall, T. Voigt, Contiki - a lightweight and flexible operating system for tiny networked sensors, in 29th Annual IEEE International Conference on Local Computer Networks, 2004, 455–462.