A Predictive Approach for Managing Network Port Resources of Service Providers

Feyzullah Alim Kalyoncu
Bogazici University,
Istanbul, Turkey, 34342
feyzullah.kalyoncu@boun.edu.tr

Engin Zeydan
CTTC,
Barcelona, Spain, 08860.
engin.zeydan@cttc.cat

Ahmet Yildirim
Department of Computer Engineering,
Bogazici University, Istanbul, Turkey, 34342.
ahmet.yildirim3@boun.edu.tr

***Abstract*—Internet service is highly dependent on the available network capacity (or network port resources) within the operation region of interest. Appropriate management of network port resources to provide Internet services for all subscribers (including new users) is a critical operation and need to be planned proactively for service providers. For this reason, prediction of available number of network port resources can provide proactive actions while ensuring enhanced decision making for operational units of service providers. This paper is investigating a prediction methodology to obtain the availability of number of network port resources for broadband IP network services using real-world case study of a major telecommunication infrastructure provider's data in Turkey. Our analysis is also considering the practical connection ranges of each region types (e.g. hospitals, airports, industrial, etc). Our predictor variables include both *in-network based data* (such as past and current total number of subscribers, number of available network ports) as well as *out-network based data* (such as population, online number of ads for sales and rental houses from a well-known Turkish advertisement website in the region of interest). Our results indicate that depending on the region of interest, prediction results on stable regions can yield better insights for service providers compared to booming regions of the country.**

Keywords*—*Port, resource, measurements, DSLAM, telecommunication, network.

I. Introduction

A communication system is built using any number or types of network infrastructure equipment. Those equipment types can be customer premise equipments (CPEs), copper twisted pair, fiber optic cable, Digital Subscriber Line Access Multiplexer (DSLAM) port, distribution point, main distribution frame and splitter [1]. Service Providers are deploying broadband IP networks nationwide to provide subscribers better and scalable Internet services. A service provider can manage or serve many regions of the country to provide nationwide coverage. At the same time, service providers need to provide high user satisfactions while delivering their subscriber facing services. Moreover, those services provided by service providers should support carrier grade high availability and reliability requirements. Operators can provide high and nationwide service availability with proper network design and proactive actions within the network.

To increase the subscriber satisfaction rates all the times and comply with Service Level Agreements (SLAs)/regulations of regulatory bodies, DSLAM network port resources should be managed appropriately and need to be up and available when subscriber demand arises. As a matter of fact, several in-network as well as out-network based factors including total number of subscribers, number of available/empty ports, population, availability of legacy and new infrastructure, number of online advertisements etc. can affect the availability of Internet service provided by service providers in the region of interest. Generally, ensuring instant availability of network resources will increase the quality-of-service (QoS) of subscribers as well as their perception on service providers. Depending on service requirements and demands of subscribers, a service provider can perform forecasts to predict the likelihood of additional or excessively needed resources in certain regions of the country. For example, an accurate prediction of available network resources such as DSLAM network ports is becoming an increasingly important service quality indicator for operators. The predictive analysis can increase subscriber satisfactions by accurate adjustments of number of available network port resources in the regions of interest. Based on the accuracy of such predictions, service providers can boost/decrease the capacity of the existing telecommunication infrastructure and can mobilize their existing workforce into the region of interest as a proactive decision maker.

There are different studies on predicting the network resource availability in different regions of the world [1], [2], [3], [4], [5]. The authors in [2] are studying availability of Internet service and propose a method that predicts the availability of IP-VPN end-to-end service in Taiwan. The authors in [1] have proposed a multivariate recurrent neural network model to predict the number of failures in broadband networks of a leading Croatian telecom provider. For example, the authors in [6] have demonstrated up to 30% energy savings at the Points of Presences (PoPs) of a nationwide Italian Internet service provider. Prediction of number of user equipments (UEs) during the time of the day with Bayesian Neural Networks using a real-world Base Station (BS) dataset in Turkey is studied in [3]. Clustering [7] and factor analysis [8] approaches are also utilized using real network datasets to observe the behaviour of network resources within Mobile Network Operators (MNOs). With respect to DSLAM based research, different DSLAM metrics are monitored in [9] to provide high quality IPTV delivery service. Broadband network markers including DSLAM port mismatches are used with fault detection and localization algorithms to predict failures in [10]. On the other hand, extensive analysis of real world datasets can also provide valuable savings for service providers [3], [6]. Note that accurate prediction of expected number of network DSLAM ports that will be needed in the future in a region of interest can provide sufficient time for proactive action and adequate preparation interval. However to the best of our knowledge, DSLAM port resource management problem using real world data has not attracted much attention in the literature.

This paper presents a real-world case study using the data of a major telecommunication infrastructure provider in Turkey. Our focus in this paper is on predicting the number of required network port resources in DSLAMs for operators that are operating nationwide in different regions of the country. Our analysis is also considering the practical connection ranges of each region type (e.g. hospitals, airports, industrial, etc). The utilized prediction model uses many factors including *in-network* and *out-network* based information and forms a connection range graph for each region to predict the total number of network ports in advance. This prediction will not only help operators to extract the potentially required number of additional (or fewer) port resources that each operator needs in the region of interest, but will also provide higher subscriber satisfactions. Our simulation results indicate that especially in relatively more stable regions of the country, prediction of additional resources yields better insights for service providers compared to booming regions of the country due to expected behaviour of constant port additions of service providers in booming regions. Focusing on both *in-network* and *out-network* (online real estate-based) data, we also identify the impact of different factors on predicting the number of available DSLAMs port resources for each region of interest.

The rest of the paper is organized as follows: Section II is presenting the system models and concepts as well the architecture. Section III gives the total number of port predictor in the given regions. Section IV is presenting the numerical result comparisons and finally Section V gives the conclusions.

II. System Model and Architecture

The system architecture exemplary of DSLAM, central office (CO), Port Resource Manager and Proposed Predictor that is studied in this paper is given in Fig. 1. DSLAM is connected to each CPE in the given region of interest and contains multiple ports where each CPE is connected to via subscriber lines. The main aim is to transport the subscriber's traffic from CPE into the nearest CO. This can be achieved via MPLS connections where either SDH or radio links can be utilized depending on the geographical conditions, distance or capacity. A region can represent one or more neighborhoods, university, industrial areas, etc. as given in Table I. Each region can be connected via legacy infrastructure, such as twisted copper pair or relatively better infrastructure such as fiber. In region I (sub-urban type) of Fig. 1, there exists a CO that is connected with DSLAM and can serve one or more residential/commercial households. Each CO is communicating with the port resource manager which runs our proposed predictive algorithm to predict the available number of port resources at each DSLAMs.

We assume that in our system there are R regions with regions set $\mathcal{R} = \{r_1, r_2, \ldots, r_R\}$. Each region $r_i \in \mathcal{R}$ has M_j DSLAMs from the DSLAM set $\mathcal{M}^{r_i} = \{m_1^{r_i}, m_2^{r_i}, \ldots, m_{M_j}^{r_i}\}$ and $A_{m_i}^{r_j}(t)$ number of subscribers at time $t \in \mathcal{T} = \{\ldots, t_0 - 2, t_0 - 1, t_0, t_0 + 1, t_0 + 2 \ldots\}$ in weeks. Each DSLAM $m_j^{r_i} \in \mathcal{M}^{r_i}$ at region r_i has $P_{m_j}^{r_i}(t)$ total number of ports. Moreover, each DSLAM $m_j^{r_i} \in \mathcal{M}^{r_i}$ is assumed to have N_{m_j} neighbor DSLAMs within a connection range $C_{m_j}^{r_i}$ with neighbour set $\mathcal{N}_{m_j^{r_i}} = \{n_1^{r_i}, n_2^{r_i}, \ldots, n_{N_{m_j}}^{r_i}\}$.

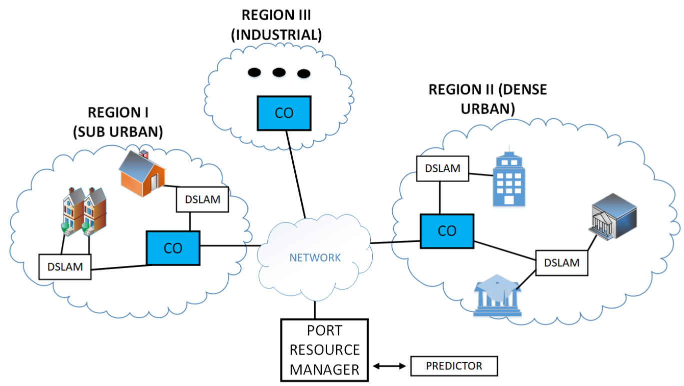

Fig. 1: System architecture illustrating DSLAM, CO, port resource manager and proposed predictor.

TABLE I: Region types and connection range in meters for a DSLAM $m_j^{r_i} \in \mathcal{M}^{r_i}$

Region Type	Connection Range $C_{m_j}^{r_i}$ (m)
AIRPORT	100
DENSE_URBAN	190
HOSPITAL	241
HOT SPOT	194
INDUSTRIAL	197
RURAL	225
SEASONAL AREAS	242
SUBURBAN	277
UNIVERSITY	221
URBAN	230

The port resource manager monitors each region of interest (urban, sub-urban, university, hospital, etc) for feature information and changes that occur in each region. It is also connected to the proposed predictor. The information acquired by the port resource manager is related to *infrastructure topology, number of available ports per network element, total number of subscribers* and *total number of port capacity in the region of interest for the considered date and time*. In addition, the port resource manager is also collecting the relevant information from online retail website for information related to *number of rental and sale houses in the region of interest*. Predictor is used to predict the potential DSLAM port resource availability k-weeks in advance. This information can be used to reduce the service issues before subscriber's complaints emerge in the field. The changing characteristics of the network such as exceeding subscriber numbers beyond regional available capacity can potentially result in network availability and performance issues for operators. For example, in the event that there is a boom of new housing and building advertisements in a given region of interest, the port resource demand per DSLAM is also expected to experience an increase. During such circumstances, operators need to scale up their available network port resources in the booming regions of interest. This can sometimes be apparent in some booming regions, however such predictions of number of available network port resource demands or rate of change in network port utilization ratios either in decreasing or increasing trends can be hard in relatively stable regions. This is true especially for old-towns and historical regions of the city where operators or service providers have already established an infrastructure and is not considering to increase the capacity due to stable structure of number of households.

Table I shows different region types and their corresponding connection ranges in meters for a DSLAM $m_j^{r_i} \in \mathcal{M}^{r_i}$. Hence, DSLAM $m_j^{r_i}$ in each given region $r_i \in \mathcal{R}$ has different connection range limitations as given by Table I. For instance, two DSLAMs $m_j^{r_i}, m_p^{r_i} \in \mathcal{M}^{r_i}$ for $j \neq p$ in hospital regions can only form a connected graph if their distance is less than 241 m. In our analysis, the resulting total number ports becomes $\widehat{P}_{m_i}^{r_j}(t)$ at time $t \in \mathcal{T}$ in a given region r_j of DSLAM $m_i^{r_j}$ is given as:

$$\widehat{P}_{m_i}^{r_j}(t) = \sum_{i \in \mathcal{N}_{m_i^{r_j}}} P_{m_i}^{r_j}(t). \tag{1}$$

Similarly, the total number of subscribers $\widehat{A}_{m_i}^{r_j}(t)$ at time $t \in \mathcal{T}$ in a given region r_j is given as:

$$\widehat{A}_{m_i}^{r_j}(t) = \sum_{i \in \mathcal{N}_{m_i^{r_j}}} A_{m_i}^{r_j}(t). \tag{2}$$

III. Port Availability Prediction of Internet services in Turkey

In this section, a prediction model for network port resource availability problem for Internet services in Turkey is investigated. In our analysis, current time is represented by t_0 and number of ports in time $t_0 + k$ is predicted, i.e. k-weeks ahead due to the requirements of the service operator to provide ample time to prepare and take preventive action beforehand. Our **dataset** is collected from two different sources, *in-network and out-network* sources. The dataset contains the following fields which are used for input to predictive analytic algorithm: *DSLAM Name*, *DSLAM Type*, *Cabin Type*, *DSL Model*, *X-Coordinate*, *Y-Coordinate*, *Neighbourhood ID*, *City Name*, *Town Name*, *District Name*, *Quarter Name*, *Total # of subscribers*, *Region*, *Installed # of Ports*, *Date*, *Neighbor DSLAMs*, *DSL Modem* as well as online real-estate dataset for *# of Ads for House Rentals* and *# of Ads for House Sales* in each region of interest. *DSLAM Type* can be outdoor or indoor. The data ranges from 2017-07-07 to 2018-02-02. The *Population* in each region that is used in our model is collected from [11]. Out-network based information is collected from a well-known Turkish real-estate advertisement website [12] at time t_0. A general nationwide statistics of the utilized DSLAM dataset is given in Table II.

We use **X**treme **G**radient **Boost**ing (XGBoost) [13] for building a model and predicting the required network port resources at time t_0+k, i.e. k-weeks in advance of t_0. The output of the XGBoost is $\widehat{\mathbf{y}}^{r_i}(t_0 + k) = \mathbf{f}(\mathbf{x}^{r_i}(t_0, t_0 - 1, \ldots, t_0 - p))$

TABLE II: Nationwide statistics of utilized DSLAM dataset in Turkey.

# of measurements	4,207,544
# of cities	81
# of ports	13,469,832
# of DSLAMs	122,143
# of neighborhoods	12,443
# of regions	98,028
Total # of subscribers	9,413,139
# of distinct DSL modems	26

where **f(.)** is the function that is matching XGBoost algorithm, p is the duration of the historical dataset, $\mathbf{x}^{r_i}$ is the vector with values of the following utilized factors: *Total # of subscribers*, *Total # of Ports*, *Population*, *# of Ads for House Rentals at* t_0, *# of Ads for House Sales at* t_0, *Date* in the region of interest. $\widehat{\mathbf{y}}^{r_i}(t_0 + k)$ denotes the vector of predicted *Total # of Ports* at time $t_0 + k$ of region $r_i \in \mathcal{R}$. We denote the train set as ($\mathbf{x}_{train}^{r_i}$, $\mathbf{y}_{train}^{r_i}$) and test set as ($\mathbf{x}_{test}^{r_i}$, $\mathbf{y}_{test}^{r_i}$). We use Root-Mean-Square Error (RMSE) metric for predictions of k-weeks ahead at region r_i where,

$$RMSE(r_i, k) = \sqrt{\frac{\sum_{t_0=1}^{T} (\widehat{\mathbf{y}}^{r_i}(t_0 + k) - \mathbf{y}_{test}^{r_i}(t_0 + k))^2}{T}} \tag{3}$$

Algorithm 1 shows the utilized general methodology for predicting the total number of DSLAM ports.

Algorithm 1 Total # of Network Ports Predictor

```
procedure PREDICTION(Total Number of Ports)
    for each Region r_i ∈ R do
        for each DSLAM m_j^{r_i} ∈ M^{r_i} do
            Map DSLAM m_j^{r_i} into region-r_i ∈ R.
            Calculate neighbor-list N_j^{r_i} using Table I.
        end for
        Calculate P̂_{m_i}^{r_j}(t) using (1), ∀t ∈ T.
        Calculate Â_{m_i}^{r_j}(t), using (2), ∀t ∈ T.
        Train a XGBoost model over (x_train^{r_i}, y_train^{r_i}).
        Predict ŷ_test^{r_i} = f(x_test^{r_i}).
        Compute RMSE(r_i) using (3).
    end for
    Obtain average RMSE(r_i) ∀r_i ∈ R.
end procedure
```

IV. Numerical Results

Without loss of generality, our focus is on outdoor DSLAM types. To obtain prediction accuracy, we use the first 6 months of data for learning and the remaining data for model testing purposes. Our data contains 7012459 sample points. We use 70% for training and the remaining 30% points for validation purposes. Each one week advance value is predicted using $p + 1 = 26$ previous time-series values. Once learning is accomplished via training, prediction is performed for one-month ahead. Therefore, we set $k = 4$ weeks for prediction time. Note that our evaluations and prediction values are based on observations only, therefore no proactive measures in real-world network infrastructure are taken after predictions

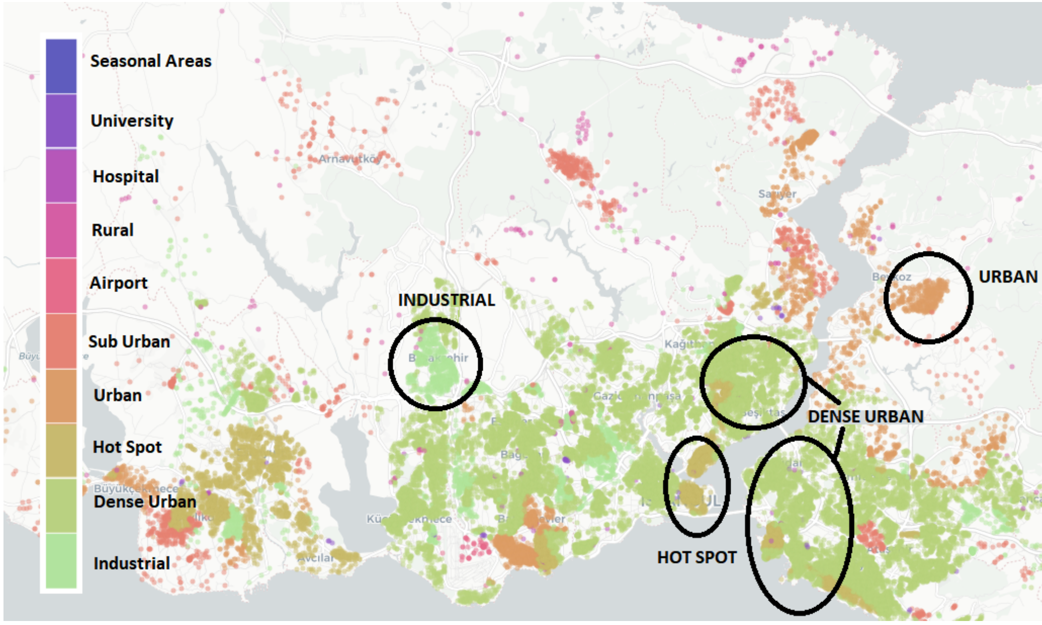

Fig. 2: An example of DSLAM distribution over Istanbul city of Turkey after aggregation of DSLAMs.

TABLE III: Factors and their importance values calculated by XGBoost algorithm.

Importance	Factor
1111	Tot. Subs. #, t_0
536	Tot. Subs. #, $t_0 - 1$
410	Installed # of Ports, $t_0 - 25$
379	2013
322	Tot. Subs. #, $t_0 - 25$
291	Ad # of House Rentals
277	Ad # of House Sales
272	Tot. Subs. #, $t_0 - 2$
...	...
4	Quarter Name
3	Town Name

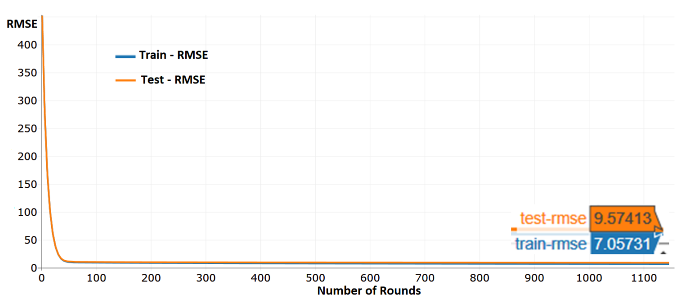

Fig. 3: Average RMSE values of XGBoost predictor over iterations for the utilized training and test dataset.

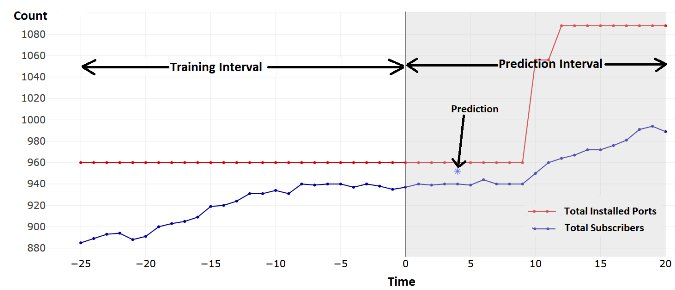

Fig. 4: Stable Region Example: Port number variation versus time.

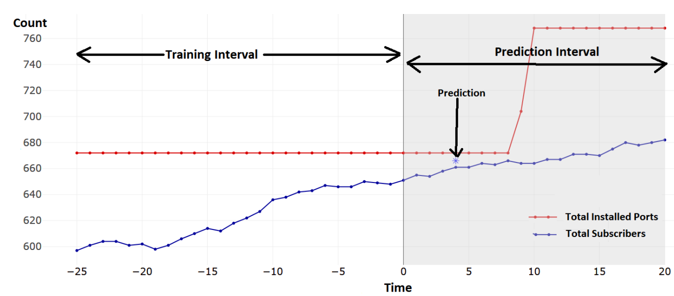

Fig. 5: Booming Region Example: Port number variation versus time.

are obtained. Fig. 2 shows an example of DSLAM heatmap distribution over different regions of city of Istanbul in Turkey. In Fig. 2, all corresponding DSLAMs that are within the given connection range for the regions of interest are aggregated.

Table III shows the top and bottom factors sorted according to their importance values utilized by XGBoost algorithm after network port predictions. The importance factor shows how important/influential the feature is for predicting the current output. In this table, t_0 represents $2017-12-29$ and $t_0 - k$ represents k weeks before t_0. The importance value represents the number of times a feature is used to split the data across all trees. We can observe that the highest importance factor for XGBoost algorithm is *total # of subscribers,* t_0 with importance value of 1111. We can also observe the importance of *Installed # of Ports* as well as *Ad # of House Rentals* and *Ad # of House Sales* at top of the Table III. To view how learning process is executed, we also show RMSE over 1000 iterations in Fig. 3. At 1100-th iterations, RMSE values with test and training dataset are around 9.57 and 7.05 respectively.

Variations of the total number of installed ports and total number of subscribers are depicted in Fig. 4 and Fig 5. In these examples, Fig. 4 represents one of the region in Halkali district (a relatively stable dense urban region in Istanbul) whereas Fig 5 is an example of booming dense urban region in Mahmutbey district. These figures also indicate the predicted values at $t_0 + 4$ week represented as blue colored dot inside prediction interval. Fig. 6 shows the actual and predicted values of the total number of ports in all given regions of Turkey. Using the results of Fig. 6, mean predicted values are computed to be within 1.9% of the actual mean values. We can observe from Figs. 4 (stable region) and Fig 5 (booming region) that based on in-network and out-network based historical data, the predicted values signify a need for major port augmentations into the existing infrastructure. However, due to not taking into account the connected DSLAMs' port graph in the considered region as well as the non-predictive approach of the considered historical data, no proactive actions have been taken. Note that during this one-month period, most of the incoming subscribers can experience a port non-availability problem. This can create a huge subscriber dissatisfaction.

During predictive interval of Fig. 4 and Fig 5, a major increase in number of subscribers has been observed after the total port capacity are increased at weeks $t_0 + 9$ and $t_0 + 8$ respectively. We can also observe a major advantage of predictive modeling over stable regions compared to booming regions. In stable regions of the city, no punctual proactive measures are taken by service provider due to non-availability of appropriate datasets such as out-network based information. However, in booming regions constant port additions by service providers is expected and normal. Therefore taking appropriate actions, e.g. addition/extraction of new DSLAM ports, takes longer time in stable regions compared to booming regions (due to unexpected increasing trends of subscriber numbers in stable regions).

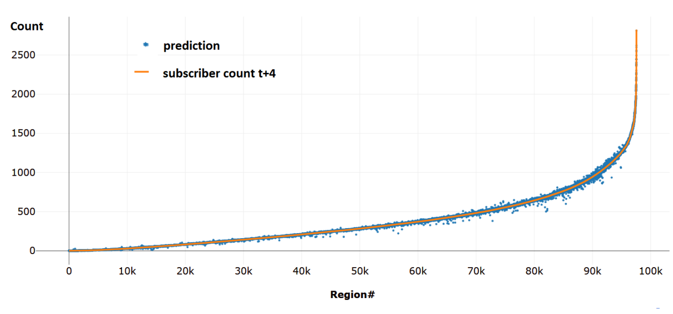

Fig. 6: Predicted versus actual number of ports over all regions of interest.

V. Conclusions

In this paper, we investigated a predictive approach to network port resource availability problem of service providers. In our analysis, we first combined all DSLAMs in a given region under the same connectivity region based on the network connectivity requirements of each region. Later, we used XGBoost algorithm to predict the expected number of ports one-month in advance for each region of interest using both in-network and out-network based data. Our results indicate that especially in relatively stable regions of the country, prediction of additional resources yields better insights for service providers compared to booming regions of the country. This is due to fact that during predictive interval, a major increase in number of subscribers of stable region has been observed after the total port capacity are increased.

Acknowledgements

This work was partially funded by Spanish MINECO grant TEC2017-88373-R (5G-REFINE) and by Generalitat de Catalunya grant 2017 SGR 1195.

References

[1] Ž. Deljac, M. Randić, and G. Krčelić, "A multivariate approach to predicting quantity of failures in broadband networks based on a recurrent neural network," *Journal of network and systems management*, vol. 24, no. 1, pp. 189–221, 2016.

[2] J.-W. Tsai, C.-S. Tang, C.-L. Lin, C.-Y. Twu, and S.-W. Yu, "Availability analysis of a broadband ip network in taiwan," in *Advanced Communication Technology, 2004. The 6th International Conference on*, vol. 2, pp. 923–926, IEEE, 2004.

[3] O. Narmanlioglu, E. Zeydan, M. Kandemir, and T. Kranda, "Prediction of active ue number with bayesian neural networks for self-organizing lte networks," in *Network of the Future (NOF), 2017 8th International Conference on th*, pp. 73–78, IEEE, 2017.

[4] L. Gupta, M. Samaka, R. Jain, A. Erbad, D. Bhamare, and C. Metz, "Colap: a predictive framework for service function chain placement in a multi-cloud environment," in *Computing and Communication Workshop and Conference (CCWC), 2017 IEEE 7th Annual*, pp. 1–9, IEEE, 2017.

[5] A. Mohamed, O. Onireti, M. A. Imran, A. Imran, and R. Tafazolli, "Predictive and core-network efficient rrc signalling for active state handover in rans with control/data separation," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1423–1436, 2017.

[6] E. Bonetto, A. Finamore, M. Mellia, and R. Fiandra, "Energy efficiency in access and aggregation networks: From current traffic to potential savings," *Computer Networks*, vol. 65, pp. 151–166, 2014.

[7] F. Kalyoncu, E. Zeydan, and I. O. Yigit, "A data analysis methodology for obtaining network slices towards 5g cellular networks," in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, pp. 1–5, IEEE, 2018.

[8] F. Kalyoncu, E. Zeydan, and A. Yildirim, "An experimental study of factor analysis over cellular network data," in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, pp. 1–5, IEEE, 2018.

[9] R. de Fréin, C. Olariu, Y. Song, R. Brennan, P. McDonagh, A. Hava, C. Thorpe, J. Murphy, L. Murphy, and P. French, "Integration of qos metrics, rules and semantic uplift for advanced iptv monitoring," *Journal of Network and Systems Management*, vol. 23, no. 3, pp. 673–708, 2015.

[10] L. Gupta, M. Samaka, R. Jain, A. Erbad, D. Bhamare, and H. A. Chan, "Fault and performance management in multi-cloud based nfv using shallow and deep predictive structures," in *2017 26th International Conference on Computer Communication and Networks (ICCCN)*, pp. 1–8, IEEE, 2017.

[11] "TSI: Turkish Statistical Institute." http://www.turkstat.gov.tr/, 2018. [Online; accessed 26-June-2018].

[12] "Sahibinden.com: Turkey's largest real-estate advertisement website." www.sahibinden.com.tr, 2018. [Online; accessed 26-June-2018].

[13] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, ACM, 2016.