Unlicensed Spectrum Forecasting: An Interference Umbrella Based on Channel Analysis and Machine Learning

Kostas Chounos, Panagiotis Karamichailidis, Nikos Makris, and Thanasis Korakis

***Abstract*—In this work a novel framework for predicting future interference levels for IEEE 802.11 networks is developed and experimentally evaluated. At the heart of the framework lies a modelling mechanism which is able to estimate and determine in real-time, the over-the-air performance that each network user will receive over an IEEE 802.11 link, when considering and combining multiple wireless metrics, without requiring a network association (Wi-Fi AP-STA) to be performed. On top of the solution, a Machine Learning approach is integrated, in order to project the real-time predictions to long-term predictions in the future (2-hour interval). Additionally, the framework applies a self-correcting mechanism for the predictions, by extracting short-term predictions and accurate throughput calculations, when the current channel conditions largely differ from the long-term predictions. The proposed framework covers comprehensively the cases of interference created by either 802.11 or non 802.11 devices, which may occur at the target or at any overlapping wireless channel. Finally, extensive testbed experimentation proves the framework's proper functionality and accuracy, under the cases of both Indoor (controlled interference) and Outdoor (uncontrolled massive interference) environments.**

***Index Terms*—Performance prediction, interference prediction, machine learning, 5G networks.**

I. Introduction

THE ever-increasing demand for continuous wireless connectivity over the past years, drastically formats the landscape of future communication systems. The constantly increasing wireless data demands have gradually rendered the existing infrastructures inadequate to cope with the increasing loads. Furthermore, as the demand for ultra low-latency and Gbps communications is constantly growing with the advent of 5G networks, proper ways for overcoming the spectrum crunch phenomenon are being investigated. Two approaches are considered: 1) the usage of new spectrum bands (e.g. the cm- and mm-Wave bands), and 2) the efficient use of existing sub-6GHz technologies, with advanced spectrum coordination techniques. The prevailing approach so far, urges the transition from macro cell architectures to the deployment of small and low range base stations that complement the coverage of macro-cells. Therefore, a combination of new-introduced and pre-existing wireless standards, can be used in order to fulfill the increased user demands in modern wireless systems.

Driven by the wide availability of Wi-Fi networks, a significant effort of the current research approaches select IEEE 802.11 (Wi-Fi) cells for dynamically serving large amount of traffic demands in 5G architectures. This can happen with either the densification of the network with heterogeneous technologies, and appropriately steering traffic to them [1], or through the integration of WiFi access in the cellular stack. An example of the latter can be found in the 4G LTE specifications, with the LTE WLAN Aggregation Adaptation Protocol (LWAAP) [2], [3]. Driven by the RAN cloudification in 5G-NR specifications, the integration of heterogeneous access is enabled through heterogeneous Distributed Units (DUs) [4]. Such DUs can support different technologies, e.g. 5G-NR, LTE or non-3GPP like WiFi, and are managed by the same Central Unit (CU). In this work, we focus on such architectures, and incorporate sophisticated algorithms and processes for managing the spectrum at the access level, when considering dense deployments of such DUs.

Delving into the different characteristics of each technology, contrary to the frequency and time allocation schemes commonly applied in the licensed spectrum (cellular technologies), most of the protocols/RF devices operating in the ISM bands are designed to transmit opportunistically in a selfish manner. The aforementioned phenomenon combined with both a large number of non-3GPP (Wi-Fi) terminals and external interference sources given nowadays, frequently results to low and unpredictable performance when accessing the Wi-Fi bands. Consequently, the performance uncertainty for DUs operating at the Wi-Fi frequencies constitutes a real struggle, in the case that they need to provide guaranteed access to the network. In this context, this work aims to tackle the problem of performance uncertainty and interference prediction when operating in the Wi-Fi unlicensed spectrum, accomplished through tight coordination of protocols operating in the same frequency spectrum. Thus, proper ways for analyzing and predicting the channel quality for these unprotected frequencies are now considered

Manuscript received December 21, 2021; revised March 8, 2022; accepted May 30, 2022. Date of publication June 3, 2022; date of current version September 9, 2022. This work was supported by the European Horizon 2020 Programme for Research, Technological Development and Demonstration under Grant Agreement 101008468 (H2020 SLICES - SC). Recommended for acceptance by Prof. Fei Shen. *(Corresponding author: Kostas Chounos)*

Kostas Chounos and Panagiotis Karamichailidis are with the Department of Electrical and Computer Engineering, University of Thessaly, 38221 Volos, Filellinon, Greece (e-mail: hounos@uth.gr; karamiha@uth.gr).

Nikos Makris and Thanasis Korakis are with the Department of Electrical and Computer Engineering, University of Thessaly, 38221 Volos, Filellinon, Greece, and also with the Centre for Research and Technology Hellas (CERTH), 38334 Volos, Greece (e-mail: nimakris@uth.gr; korakis@uth.gr).

Digital Object Identifier 10.1109/TNSE.2022.3180171

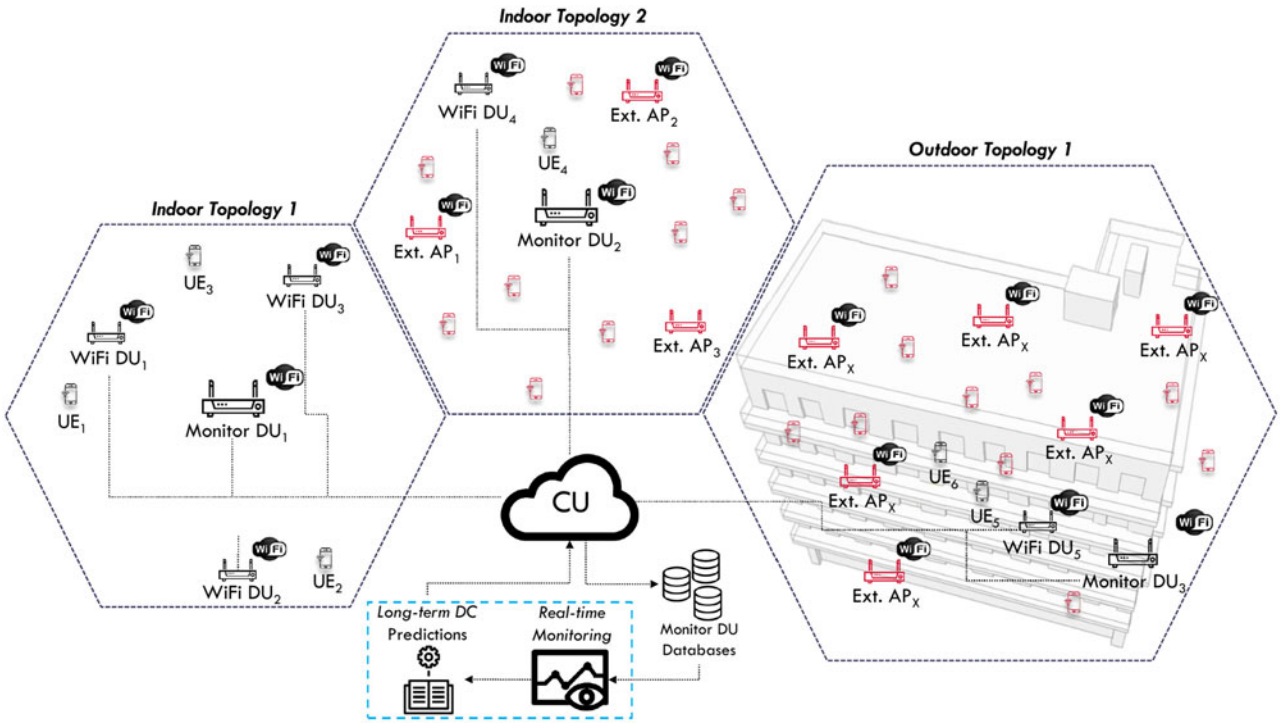


Fig. 1. Under study system architecture: dedicated monitor DUs constantly collect spectrum metrics for their area of supervision, and upload them to the CU running at the network edge. Metrics are extracted, and Machine Learning algorithms are able to predict their evolution in the future, while also estimating the expected performance for each wireless channel through their combination.

essential. Initially, the analytical channel conditions are broken-down, performed through the active collection of several wireless metrics. Through these metrics, we are able to extract accurate performance calculations for every available Wi-Fi link, either at short-term or long-term. Furthermore, Machine Learning (ML) is employed for accurately predicting the expected performance for each network, based on predicting isolated features, the combination of which can reveal the actual offered capacity. The quality prediction problem is usually formulated as a Time-series prediction problem, thus enabling the application of several ML-driven methods for predicting factors and metrics in the network that directly affect the performance over the network.

In this work, we adopt the disaggregated multi-technology base station architecture [5] and introduce a novel framework for estimating and predicting the performance of each UE in the unlicensed Wi-Fi bands. More specifically, besides the deployed CU and Wi-Fi DUs, we also introduce the entity of a Monitor DU in the considered network architecture as illustrated in Fig. 1. The presented work differentiates from similar efforts in bibliography by predicting the expected Duty Cycle ($\mathcal{DC}$) at each Wi-Fi channel, a metric that can be used alongside with other wireless analytics, to efficiently calculate the estimated capacity of each network. This is, to the best of our knowledge the first work that considers the evolution of this metric, towards providing tangible proof [6] about the estimated performance over each wireless link. ML is employed collaboratively towards predicting such metrics for the long term, used to estimate the available capacity per each link. The ML models that we use have been selected after multiple experimental evaluations of different Neural Network models, towards increasing the accuracy and lowering the execution time.

The estimation process is running on the operator side of the network, and can assist towards the selection of cells that can serve each client of the network in the most efficient manner. The overall system is experimentally driven and runs on real off-the-shelf devices, deployed in an outdoor dense urban area with entirely uncontrolled external interference settings. Contrary to our approach, prior works are mainly focusing on predictions at the client side, based on the received signal strength. The key contributions of this work can be summarized as follows:

- To dynamically estimate and predict the delivered performance (*network capacity*) that a user shall get over a target non-3GPP DU. Through the employment of advanced wireless analytics, we are able to make long and short-term accurate predictions on the utilization of each network. This is, to the best of our knowledge, the first work in which the estimation is taking place before the user association with the DU.
- To export performance estimation calculations and predictions for all Wi-Fi channels by taking into consideration analytically the interference occurred by (1) Wi-Fi terminals from overlapping channels, (2) WiFi terminals at the same channel and (3) non-IEEE 802.11 interference.
- To experimentally evaluate and showcase the gains in a real ultra-dense network setups, for several different controlled and uncontrolled interference scenarios.

II. Related Work

Performance estimation approaches for either wired or wireless networks has fostered efforts in several research works in relevant literature. As the network deployments become denser and rich in the offered technologies, their efficient coordination, planning and placement in operating spectrum is of primary importance for providing high quality connectivity services to end users. The ability to accurately calculate and predict the expected performance, may constitute a crucial factor for the proper and efficient network design and planning. Three different types for performance predictions can be extracted from the proposed research approaches [7]: 1) formula-based (e.g. [8]), 2) history-based (e.g. [9]) and 3) based on the application of machine learning (ML) (e.g. [10]). We organize the relevant works in two subsections: (1) on performance estimation, based on either formulas/historical data, and (2) based on the application of ML models.

A. Performance Estimation

This category covers works which develop mathematical formulas for modeling TCP or UDP behaviour, based on several link characteristics (rate, jitter, packet loss, round trip time etc). For history-based approaches, the estimation is calculated based on previous performance measurements (e.g. throughput calculation during a specific time period). The future throughput predictions are based on the moving averages for past values. As the wireless network can have wide fluctuations in the quality, affecting the client performance, link estimation is of crucial importance for the efficient operation of the network. Authors in [7] estimate the cellular network performance, by employing 7 different algorithms. They introduce the theory of differential entropy to estimate the lower bound on throughput predictions, and use actual cellular data for efficiently predicting the TCP throughput. In the case of non-3GPP access, estimations are harder to be made efficiently, due to the opportunistic access of the spectrum. Therefore, performance estimation is usually backed by an interference detection/estimation scheme. Authors in [11], propose a network planning scheme for Wi-Fi networks deployed in ultra dense scenarios. Their work is based on interference alignment, with their cells being separated into smaller subregions, and using the same channel allocation. The proper size of the subregions as well as the selected channel is extracted by maximizing the Degree of Freedom (DoF). However, in practical ultra dense networks and in the case of the unlicensed frequencies, there can be no assurance for the absence of external not coordinated Wi-Fi Access Points (APs) or Stations (STAs).

Similarly, in [12] a Dynamic Channel Selection (DCS) algorithm for sectorized WiFi cells is proposed. More specifically, the authors share the coordinated APs between the three "non" overlapping (1,6 and 11) channels. Under the use of Received Signal Strength Indicator (RSSI) and RF utilization metrics, they periodically create an interference map which drives the DCS. Although the mechanism measures the Wi-Fi interference created from both coordinated and uncoordinated APs, the interference occurred from channels besides 1,6 and 11 is not considered. Authors in [13] delve into interference mitigation techniques, towards enhancing the quality of the offered wireless services. Works [14]–[16] present different analytical models which focus on Wi-Fi related parameters like A-MPDU aggregation, Block ACKs and packet retransmissions. Indicatively, Nayak *et al.* [14] introduce the Virtual Speed Test (VST), an AP analysis tool for estimating the performance in both downlink and uplink TCP traffic scenarios. By passively collecting statistics for all the active TCP connections, they calculate the time needed for transmitting and acknowledging each packet. The aforementioned approach assumes that all the examined STAs are connected and actively transmit large amounts of data, in order to be identified and successfully measured.

B. ML Based Estimation

ML techniques that have been very popular lately have transformed as the prevailing approach for performance prediction especially for wireless networks. For such networks, the sophisticated pattern recognition may provide accurate predictions, driven by historical data, while adapting the estimations based on a different variables and features that might impact performance.

Their adoption is also reflected in the current efforts under the Open RAN (O-RAN) alliance [17], through the integration of ML analysis for channel quality measurements driving resource allocation in the RAN. In [18], authors describe how a data-driven architecture of a wireless network can benefit from the application of online learning, such as in the cases of load balancing, etc. In [19], authors argue about the opportunities in the cellular network for applying ML towards optimizing its behavior. Authors in [20], present an analytical survey on models and concepts based on ML that can be used for managing and organizing cellular 5G networks. In [21], authors use an edge-controller-based architecture for cellular networks and describe how the controllers can be used to run ML algorithms to predict the number of users in each base station. Finally, a use case in which these predictions are exploited by a higher-layer application to route vehicular traffic according to network Key Performance Indicators (KPIs).

Specifically to non-3GPP access, authors in [22] develop an unsupervised Neural Network to filter the detected transmission collision probability in the unlicensed spectrum. Their solution is trained online, and can precisely rectify the measurement error and estimate the number of active WiFi users from the user side. In [23], a reinforcement learning algorithm is introduced, aiming at the problem of opportunistic coexistence of unlicensed LTE and WiFi. The proposed approach in particular is based on Q-Learning, which provides a robust and model-free decision-making framework. Authors in [24], design and formulate a function for autonomous APs that estimates throughput and delay of its clients in 2.4GHz WiFi channels. The target AP can estimate throughput and delay of its clients without actually switching to each channel, but only monitoring MAC level information sent over the network. Authors in [25], [26], exploit ML for predicting wireless

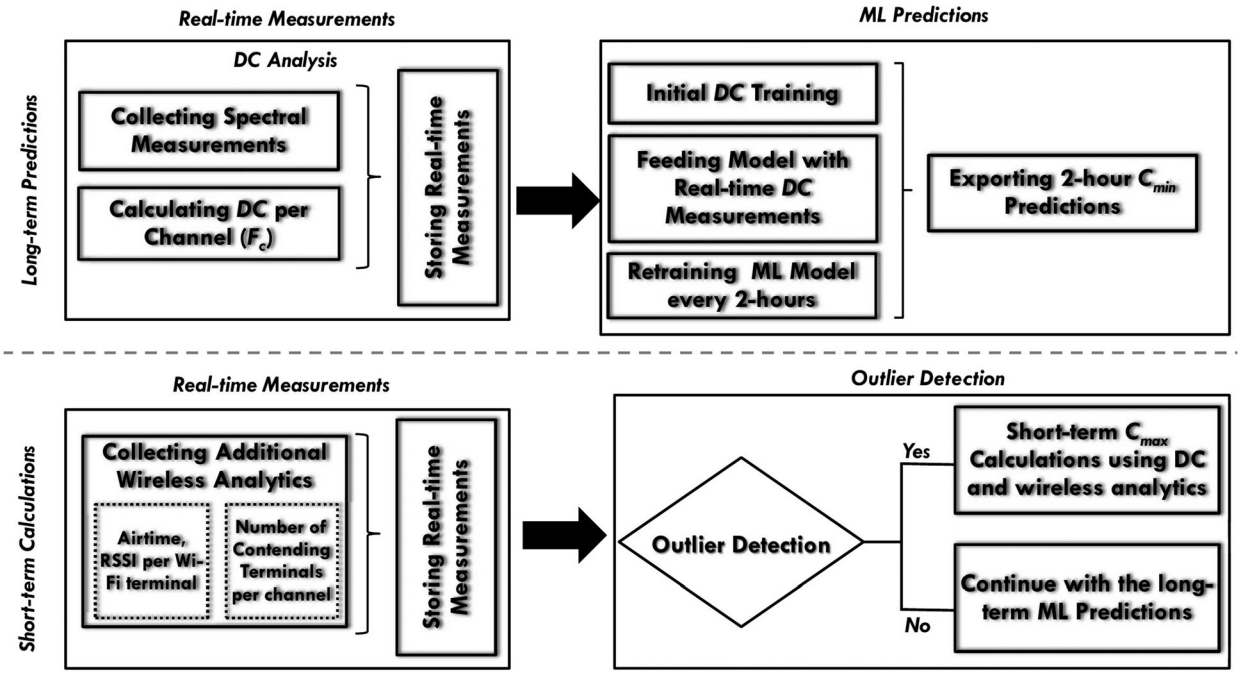


Fig. 2. Process flow for the proposed framework; the framework is able to predict accurately for the long-term, by performing DC analysis and using our ML model. The long-term predictions are validated through extended analytics, and if found invalid are corrected with the short-term predictions.

channel quality. In [25], a deep learning scheme is applied to predict wireless interference, showcasing gains in Signal to Noise Ratio (SNR) monitored at the network's clients. Struye *et al.* in [26] propose a Neural Network-based Interference prediction scheme. More specifically, the authors try to deal with the unexpected presence of interference existing at city environments. They employ a Gated Recurrent Neural Network (RNN) towards forecasting the RSSI values for each deployed AP at different times of the day.

C. Contribution With Respect to State-of-The-Art

In this work, we develop an analytical framework for estimating the network capacity of non-3GPP Wi-Fi networks, using a wide set of wireless analytics. We differentiate from the previous works by focusing on the prediction of the Duty Cycle ($\mathcal{DC}$) metric, contrary to prior approaches focusing on common values such as the RSSI, which in turn can be used to efficiently estimate the available network capacity that can be used by each network client. The $\mathcal{DC}$ metric can generalize to any type of network which operates at any frequency range, by showing the measured energy values over time (spectral scan duration) in defined chunks of spectrum. Thus, it can be applied to networks using other access mechanisms and different physical access schemes other than the ones used in Wi-Fi (CSMA/CA and OFDM/OFDMA). We apply the scheme in a novel disaggregated 5G Cloud-RAN, based on our prior contributions in [27]. The ML approach is based on a Gated Recurrent Unit (GRU) Neural Network, which has been selected over other approaches after extended testing and evaluation of the training time needed and achieved prediction accuracy. In the following section we detail the proposed framework and ML model that we use.

III. Framework

In this work we design, develop and experimentally evaluate a hybrid framework for estimating, predicting and calculating channel performance for each client operating in the unlicensed Wi-Fi bands in real-time. Our framework is able to conclude on the estimated channel performance based on a set of wireless metrics on the expected performance of a wireless client before attaching to a Wi-Fi DU for the short-term. Towards expanding our solution for the time, and by keeping accurate backlogs of historic measurements for a specific geographical area, we are able to extend our solution towards accurately making long-term predictions. By actively gathering several wireless metrics which are obtained from commercial off-the-shelf (COTS) devices, the proposed framework is able to predict the transmission opportunities per channel for the next 2-hours, with a minute-based resolution through ML. When the ongoing spectral conditions largely differ from the long-term predicted, an additional history-based short-term mechanism is activated and calculates the exact performance which can be currently achieved. Both mechanisms run in parallel for collecting metrics, and make conclusions only when the framework decides that they should run. Fig. 2 shows how the different components interact with each other. In the subsections below, we detail how each of the components works. It is worth to be noted, that the proposed mechanism does not require in any case Wi-Fi DU - UE associations. The next sections describe how the framework is organized: Sections III-A and III-B show the internals of the long term prediction model, while section III-C and III-D describe how we employ additional analytics and make short-term estimations on throughput.

DC_CH1	DC_CH2	DC_CH3	DC_CH4	DC_CH5	DC_CH6	DC_CH7	DC_CH8	DC_CH9	DC_CH10	DC_CH11	DC_CH12	dateandtime
14.1117	12.3165	8.7763	9.59709	12.4723	19.9179	29.9494	37.5611	42.8856	45.3956	45.4768	39.5671	2021-12-03 09:21:04
10.9772	9.65471	7.11004	8.94888	12.3979	19.9946	30.075	37.8616	43.1091	45.0226	44.4397	38.8196	2021-12-03 09:21:15
10.9566	9.47771	7.00365	7.92587	10.7636	18.1292	27.2835	35.4563	41.0626	42.235	41.5156	36.1047	2021-12-03 09:21:26
25.9588	22.2022	13.8046	13.9273	18.6187	25.8488	35.1426	42.0583	45.5696	46.8902	46.7794	39.6791	2021-12-03 09:21:37
23.735	20.3826	14.4413	11.1406	9.58551	12.7558	17.8521	22.9787	26.9172	29.1003	29.7934	26.1921	2021-12-03 09:21:48

Fig. 3. $\mathcal{DC}(\%)$ measurements for channels (1-12) of the 2.4GHz Wi-Fi band, which are continuously collected from Monitor DUs. The aforementioned Data Sets - Measurements, are used in the proposed framework, both at short-term calculations / estimations, as well as at the long-term ML predictions.

A. Data Set Description

In the context of this work, we use part of a data set[1] which uninterruptedly stores measurements since 2016. Similarly with [26], we take advantage of the spectral measurements which can be revealed through the application of Energy Detection on commercial Wi-Fi chipsets. Indicatively, we have deployed a Monitor DU on the top of an office building (3 rd floor) in the densely populated city center of Volos, Greece. In particular, the TP-link Archer C7 V2 [30] router is engaged to run as a spectrum analyzer, monitoring of all the sub-6GHz IEEE 802.11(a/ac/b/g/n) wireless standards. In contrast with prior works, in which the RSSI is used as a key metric, we extract the Duty Cycle ($\mathcal{DC}$) of each scanned channel based on the spectral measurements. For each Wi-Fi channel ($\mathcal{F}_c$), we collect multiple spectral samples ($\mathcal{N}_S$) and export the $\mathcal{DC}$ percentage utilization as follows:

$$\mathcal{DC}_{\mathcal{F}_c} = \frac{1}{\mathcal{N}_S} \sum_{S=1}^{\mathcal{N}_S} on(\mathcal{P}(\mathcal{S}, \mathcal{F}_c), \mathcal{P}_{\mathcal{TH}}), \tag{1}$$

More specifically, we use $\mathcal{P}(\mathcal{S}, \mathcal{F}_c)$ to denote the power of each Spectral Sample ($\mathcal{S}$) that has been collected on the central frequency $\mathcal{F}_c$. Thereupon, each Spectral Sample's power is compared with a power threshold ($\mathcal{P}_{TH}$) and consider it "on" if found higher than that. In order to set the $\mathcal{P}_{TH}$, the proposed framework uses the Energy Detection (ED) and Clear Channel Assessment (CCA) thresholds, as defined per each IEEE 802.11 standard version. Additionally, we collect 200 $\mathcal{N}_S$ per $\mathcal{F}_c$ for increasing the precision of the detection process. It is worth to be noted that the spectral samples are captured in 20 MHz width and with 56-FFT bin resolution. There are multiple spectral sample configurations which are supported from the wireless adapters (up to 80 MHz with 256 bins). However, we employ this configuration at the lowest channel settings, as it may be multiplied for higher channel widths without any information loss. Through the proper adjustment of the CCA/ED thresholds, the $\mathcal{DC}$ information is calculated for the primary and secondary(ies) channels, and useful utilization and performance information for 20-160 MHz wide channels can be retrieved. Based on the applied configurations, the spectral scan for all channels in both 2.4GHz and 5GHz is calculated within ~240 ms. The detection accuracy of the aforementioned approach has been proven analytically in [6]. Furthermore, we trigger each monitor DU to obtain spectral measurements and calculate the $\mathcal{DC}$ utilization values for all 20 MHz Wi-Fi channels every 10 seconds. While the aforementioned procedure is uninterruptedly executed from May 2016 until today, we now hold a data set of ~16 million $\mathcal{DC}$ values for each frequency. The $\mathcal{DC}$ Data Set structure for a Monitor DU, as well as an indicative part of its measurements, are depicted in Fig. 3.

As the $\mathcal{DC}$ notion is calculated through multiple samples which are collected during the spectral scan procedure, some interesting facts can be summarized as follows. Apart from the active interference intensity, that can be extracted solely through the RSSI metric, the $\mathcal{DC}$ offers the ability to quantify its impact over time as well. In such way, we are able to accurately calculate the achieved performance for any Wi-Fi link, at the areas of supervision. Additionally, both the amount of interference occurred from IEEE 802.11, as well as for non-IEEE 802.11 devices (e.g. microwave ovens, baby monitors, wireless cameras, Bluetooth and Zigbee devices etc) are included in the $\mathcal{DC}$ metric. Any form of energy emission / transmitted signal, which takes place at the Wi-Fi bands, can be detected and quantified accurately through the utilization of $\mathcal{DC}$ metric [6]. Furthermore, medium and frequency access schemes (e.g. CSMA/CA, OFDM, OFDMA), do not affect $\mathcal{DC}$ calculation, as it is extracted by raw energy of spectral samples which are taken over time.

By taking into account (1) and all the remarks stated above, the $\mathcal{DC}$ metric may also represent the minimum free transmission opportunities, which are given at each $\mathcal{F}_c$:

$$\mathcal{C}_{min} = (1 - \mathcal{DC}_{\mathcal{F}_c}) * \mathcal{T}_{\mathcal{MCS}}, \tag{2}$$

where $\mathcal{T}_{\mathcal{MCS}}$ is the theoretical throughput corresponding to each case Modulation and Coding Scheme (MCS), for either UDP or TCP traffic streams [31]. Finally, by gathering additional wireless metrics as described in the following Sections, this work tackles the problem of the performance estimation uncertainty for IEEE 802.11 systems and exports the maximum transmission opportunities ($\mathcal{C}_{max}$) which can be achieved for each channel.

B. Machine Learning $\mathcal{DC}$ Training and Prediction

In the context of this work, we additionally develop Machine Learning (ML) models to make long-term (2-hour) predictions of the $\mathcal{DC}$ utilization for all Wi-Fi channels. Specifically, we make a prediction for the $\mathcal{DC}$ utilization every minute into the future, based on historic measurements. Thus,

[1] In this work, all the Data Sets - Measurements, are collected from authors research group (NITlab) [28] and in the University of Thessaly's facilities. Additionally, the Duty Cycle measurements collected at 2.4GHz Wi-Fi band from Monitor DU3 (2021), can be found in [29].

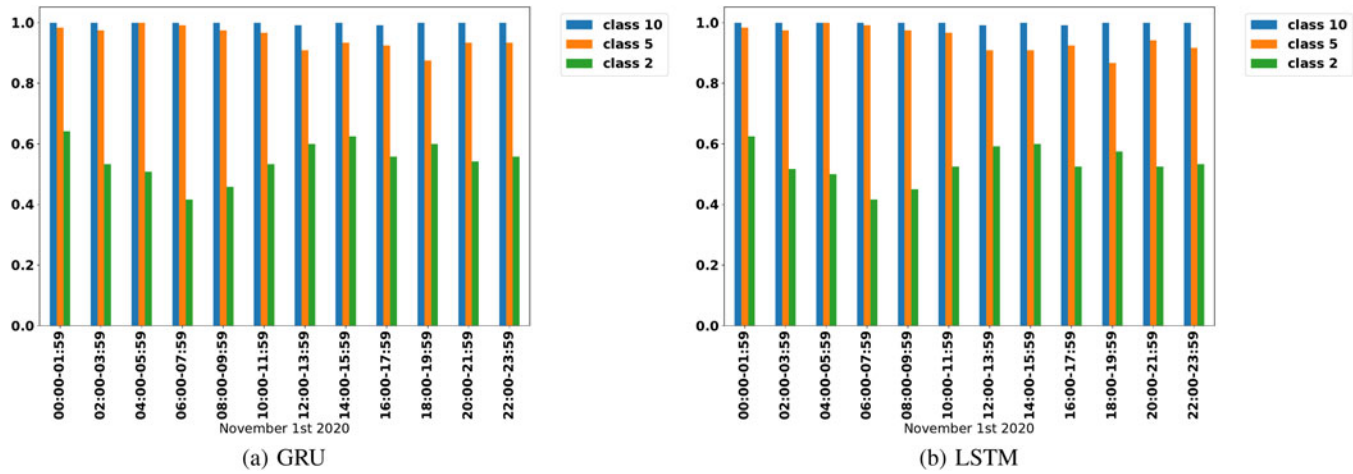


Fig. 4. $\mathcal{DC}$ Class Prediction Accuracy (per-minute).

a detailed analysis for the expected $\mathcal{C}_{min}$ of every Wi-Fi terminal is outputted, by jointly taking into account, the frequently updated RSSI values in the central database. Since the problem we tackle is a time series prediction, we use Recurrent Neural Networks (RNNs). This type of networks have been proven to be appropriate for modeling sequence data [32], [33], due to the fact that they can recognize temporal patterns by remembering aspects of past inputs in their internal memory.

In particular, we have tested Gated Recurrent Unit (GRU) [34] and Long short-term memory (LSTM) [35] models, which are implementations of the RNN concept. Both deployed models contain two stacked GRU/LSTM layers of 500 units each, followed by a fully connected layer which reduces the outputs to a single value. Additionally, they are trained for 100 epochs and by using the adam optimizer [36]. Furthermore, early stopping with patience argument [37] was used, in order to make model training procedure even faster. In such way, when the monitored metric stops improving, the training process is being terminated and based on the defined patience given.

As input, we feed the models with a part of the $\mathcal{DC}$ dataset described in Section III-A. There, an uninterrupted monitoring of the spectrum conditions, is taking place from May 2016 and until today. As it is analytically described, through the appropriate use of (1), we are able to calculate the $\mathcal{DC}$ utilization for each Wi-Fi channel every 10 sec. In this work, we use the most recent $\mathcal{DC}$ measurements which are from January until October 2020. The input data are split in 70-30 ratio for training and validating the models accordingly. Furthermore, the measurements are normalized between 0 and 1. Since the target prediction is for two hours in the future, the model is retrained every two hours to keep up with the real-time collected $\mathcal{DC}$ measurements. The machine used for (re)training our models is using an Intel Core i7-8700 CPU at 3.20GHz with 16GB RAM. The initial training process lasts approximately 20 minutes for GRU and 45 minutes for LSTM. Due to the fact that we use an early stopping with patience of 25, while monitoring the validation loss, the GRU model stops training at 43 epochs, compared to LSTM which terminates training at 100 epochs. It is also noted, that the retraining procedure lasts less than a minute for both models. Additionally, smaller patience for the LSTM model could be given, which may lead to faster training. However, it was deemed correct to compare these models under exact same configurations. Several configurations (layer units, epochs, patience etc) were extensively tested for the developed ML models. However, the aforementioned configurations were selected as they efficiently fulfill the dynamic requirements of the proposed framework, with the minimal overhead.

However, in real world uncontrolled spectrum environments, the channel utilization and in consequence the $\mathcal{DC}$ values per second may contain great variations. This is due to the frequent arrival and departure of Wi-Fi stations, as well as in the various traffic demand profiles of each device. For this reason, the prediction of the $\mathcal{DC}$ value in a decimal place precision, would be both hard and unnecessary. In such way, for evaluating the final results, we create three classes: "class2," "class5" and "class10". These numbers represent the class size in percentages of the $\mathcal{DC}$ value. Therefore, for example in class10 we have batches of $\mathcal{DC}$ values between 0-10, 10-20 and so on until 90-100%. In the proposed model, we classify the predicted values and the actual values into these three classes and measure the accuracy of each class for both GRU and LSTM, as depicted in Fig. 4.

In order to evaluate the model, we feed it with the $\mathcal{DC}$ data from January through October 2020 and we indicatively forecast the first day of November. Within the 2-hour intervals, the GRU model achieves accuracy which ranges between 41-64%, 87-99%, 99-100% for classes 2, 5 and 10 respectively (Fig. 4 (a)). Respectively, for the same class sizes, LSTM achieves prediction accuracy of 40-62%, 86-98% and 99-100% (Fig. 4 (b)). Therefore, the model selected for this work is the GRU which performs slightly better for our data. Additionally, Fig. 4 confirms that great $\mathcal{DC}$ utilization variations are given at those uncoordinated outdoor environments. In such way, lower $\mathcal{DC}$ detection accuracy is observed for classes 2 and 5. This is happening due to the dynamics of the wireless conditions that we test our framework, as it is a dense outdoor environment with high uncontrolled external interference. This in turn causes the $\mathcal{DC}$ metric to greatly variate in smaller examined batches causing lower accuracy for the lower classes. As we consider

type	mac_address	associated_to_mac	ssid	channel	rssi	airtime	dateandtime ▲ 1
AP	C8:B3:73:21:DD	-		1	-47	2.16801	2020-10-30 15:18:51
STA	38:BA:F8:58:08:	C8:B3:73:21:DD		1	-52	0.766859	2020-10-30 15:18:52
AP	B0:AC:D2:7A:AF	-		1	-39	2.34183	2020-10-30 15:18:52
STA	D4:76:EA:26:73:	(not associated)		1	-63	0	2020-10-30 15:18:52
AP	96:B5:88:A8:F6:	-		1	-65	0.192728	2020-10-30 15:18:52

Fig. 5. Additional Wireless Analytics developed and used for increasing the accuracy of our predictions; associations per each station, RSSI and airtime percentages for each Wi-Fi device are constantly monitored and logged in a central database, enabling the creation of a rich "spectral" map for the monitored area.

higher classes, the variations in the $\mathcal{DC}$ diminish. Based on our findings, the class size 10 is selected for performing the $\mathcal{DC}$ predictions, as it offers the highest accuracy. More specifically, we keep the mean value of each predicted class for performing the long-term $\mathcal{C}_{min}$ calculations. Therefore, an error up to $\pm 5\%$ may arise for the predicted $\mathcal{C}_{min}$ value per each minute.

C. Estimated Throughput Calculation

As stated in Section III-A, the calculation of the $\mathcal{DC}$ metric can export useful information for the utilization of each Wi-Fi channel. However, this utilization percentage may contain both IEEE 802.11 and non-802.11 interference, which practically means that the estimated performance can not be characterized solely from it. A typical example could be that if an $\mathcal{F}_c$ is considered fully utilized from IEEE 802.11 stations ($\sim 100\%$ $\mathcal{DC}$), this will not result to zero transmission opportunities if a new Wi-Fi terminal would try to transmit in the same frequency. As a rule, the occupied utilization at a Wi-Fi channel can be distinctly separated in:

$$\mathcal{DC}_{\mathcal{F}_c} = \mathcal{I}nt_{(Wi-Fi)} + \mathcal{I}nt_{(Ext)} \tag{3}$$

where $\mathcal{I}nt_{(Wi-Fi)}$ defines the sum of the airtime utilization $(air_{(\%)})$ [2] for all the Wi-Fi terminals ($\mathcal{N}_{Terms}$), which transmit at the examined $\mathcal{F}_c$

$$\mathcal{I}nt_{(Wi-Fi)} = \sum_{i=1}^{N_{Terms}} (air_{i_{(\%)}}) \tag{4}$$

An indicative set of airtime measurements is depicted in Fig. 5. This metric is stored in a central database and utilized jointly with multiple other wireless analytics. In such way, the proposed framework can export the accurate performance ($\mathcal{C}_{max}$) for each Wi-Fi link, as described further below. Moreover, the $\mathcal{I}nt_{(Ext)}$ is the external interference occurred either from non-IEEE 802.11 devices, or from Wi-Fi terminals which transmit at the neighboring overlapping channels. The latter is a phenomenon happening by the design of the Wi-Fi protocol, as transmissions from neighbouring channels can be captured in a terminal but in most cases cannot be fully decoded.

[2] The airtime utilization, represents the percentage of channel time consumed from each Wi-Fi terminal (AP or STA), to transmit its frames, retransmissions, acknowledgments (ACKs), control and management frames.

$$\mathcal{I}nt_{(Ext)} = \mathcal{DC}_{\mathcal{F}_c} - \mathcal{I}nt_{(Wi-Fi)} \tag{5}$$

The substantial difference between $\mathcal{I}nt_{(Wi-Fi)}$ and $\mathcal{I}nt_{(Ext)}$ is that for the former, conventional mechanisms of the IEEE 802.11 standard as CSMA/CA are activated, and some transmission opportunities become available. In the latter, the presence of such interference exceeding the ED threshold, directly leads to failed / backed-off transmissions. For this reason, it is essential to measure the amount of interference ($\mathcal{DC}$) at each $\mathcal{F}_c$, and perceive what causes it. In this manner, we can conclude on the maximum supported capacity of a cell at a given time. Similarly with [6], we follow all the pre-defined ED/CCA standard thresholds, in order to characterize a captured signal as harmful interference, thus including it in $\mathcal{N}_{Terms}$, $\mathcal{I}nt_{(Wi-Fi)}$, $\mathcal{I}nt_{(Ext)}$ and in the airtime metric. All the calculations concerning the airtime percentages include the transmissions, re-transmissions as well as the control and management frames of each Wi-Fi device.

As the detection of $\mathcal{DC}$ metric takes $\sim$240 ms and the procedure is executed periodically every 10 seconds, we take advantage of the idle time to capture several additional useful channel analytics. The collected analytics and metrics are further discussed in this section. The overall algorithm is summarized as follows: (1) Initially, the DC analysis from spectral scan is extracted. The process is blocking meaning that the wireless devices cannot be used during this interval ($\sim$240 ms) for any transmission/reception of data. (2) The mode of the wireless device is switched to a *monitor* mode, allowing the collection of the different metrics. The metrics collected under this mode assist in the identification of possible external and IEEE 802.11 interference, from adjacent/overlapping channels. This process is lasting for 10 secs - 240 ms. In this manner, the entire analysis of the channel conditions is restricted to 10 seconds in total, ensuring that enough spectral data covering a rich number of different cases are collected. (3) From the data collected within the second interval, we calculate the maximum capacity per channel, as further defined in this section. This process is repeated periodically per minute, and runs completely at the Monitor DU entity of the network.

Regarding the collected metrics, when the wireless adapter is switched to the *monitor* mode, it acts as a packet sniffer by utilizing *airodump* [38], *tcpdump* [39] and *tshark* [40] Linux tools. By serially scanning all the Wi-Fi channels with airodump, we

are able to create every minute an additional useful spectrum map, which contains all the present AP-STA associations per channel and the RSSI for each terminal. Furthermore, the exact airtime for each Wi-Fi terminal is calculated through the use of tcpdump and tshark tools. A python script was developed in order to orchestrate the above procedure and similarly with $\mathcal{DC}$, store all the results in a database as shown in Fig. 5. In such way, the exact $\mathcal{C}_{max}$ performance which can be achieved for each $\mathcal{F}_c$, can be indicated under the proper combination of $\mathcal{DC}$ with the number of terminals and the occupied airtime. Based on IEEE's 802.11 Distributed Coordination Function (DCF), the medium should be equally distributed in all users, thus achieving service fairness [41] at each wireless channel. As $\mathcal{I}nt_{(Ext)}$ unavoidably leads to failed / backed-off transmissions, we need to further break down $\mathcal{I}nt_{(Wi-Fi)}$ percentage for exploiting the contendable transmission opportunities. In a given $\mathcal{F}_c$, the users ($\mathcal{N}_{ContTerms}$) who will be affected in terms of performance, upon the arrival of a new user will be those of whom:

$$air_{(\%)} > (1 - \mathcal{I}nt_{(Ext)})/(\mathcal{N}_{Terms} + 1) \tag{6}$$

and thus, the total percentage of the occupied contendable airtime in a channel can be calculated as follows:

$$\mathcal{C}ont_{(air)} = \sum_{j=1}^{N_{ContTerms}} (air_{j_{(\%)}}) \tag{7}$$

If we combine all equations introduced so far and by covering every possible case of fully utilized or not channels, the estimated throughput for a Wi-Fi terminal after associating at a specific $\mathcal{F}_c$ can be defined as:

$$\mathcal{C}_{max} = \frac{1 - \mathcal{DC}_{\mathcal{F}_c} + \mathcal{C}ont_{(air)}}{\mathcal{N}_{ContTerms} + 1} * T_{\mathcal{MCS}} \tag{8}$$

Despite the fact that the calculation in (8) reports the estimated throughput in a primary 20 MHz channel, the appropriate reconfiguration of the CCA/ED thresholds, extracts the results for secondary channels, in the cases of 40-160 MHz bandwidths.

D. Real-Time Outlier Detection

As described at Section III-B, accurate long-term (2-hours in a minute basis) utilization predictions can be provided, under the proper use of $\mathcal{DC}$ metric through ML. However, based on the dynamic demand profiles of the wireless users and the fluctuating dynamics of the wireless medium, several severe short-term variations may occur. These variations are considered as outliers and it is practically impossible to be predicted efficiently through ML in long-term periods [42].

In this way, we need to shield our system by dynamically detecting these short-term non periodic events. Thus, an additional rapid real-time prediction mechanism is developed and enabled when the current performance largely differs from the values which were predicted with the contributions described in Section III-B. Let us consider the topology of "Outdoor Topology 1," as illustrated in Fig. 1. Alongside with the Monitor DU_3, the Wi-Fi DU_5 is deployed and can be orchestrated from the CU. In order to successfully recognize the potential outliers, which may cause performance impairments in the short-term, we make use of the most recent $\mathcal{DC}$ measurements and the additional wireless analytics reported from Monitor DU_3. As described in the previous sections, we are able to accurately detect and calculate the impact of both IEEE 802.11 and non-802.11 interference, regardless of whether this occurs in the target $\mathcal{F}_c$, or at overlapping channels.

Initially, it is very important to clarify if any abnormal spectrum behaviour is caused from the under consideration coordinated Wi-Fi DU and its associated UE(s) ($\mathcal{N}_{CoordTerms}$), or by external interference sources. In the proposed framework, the mechanism retrieves new $\mathcal{DC}$ measurements every 10 seconds (Section III-A) and we define the current utilization $\overline{\mathcal{DC}}_{\mathcal{F}_c}(t_0)$, as the weighted moving average [43] of the last 6 measurements, which were obtained during the previous minute.

$$\overline{\mathcal{DC}}_{\mathcal{F}_c}(t_0) = \sum_{n=-1}^{-6} \mathcal{DC}_{\mathcal{F}_c}(t_n) * \mathcal{W}_{|n|} \tag{9}$$

Under the use of the weight vector $\mathcal{W}_{|n|} = [0.2857, 0.2381, 0.1905, 0.1429, 0.0952, 0.0476]$, slightly greater emphasis is given on the most recent $\mathcal{DC}$ measurements. Furthermore, the proposed framework (Section III-C) stores the additional wireless analytics one time every minute. In such way, we capitalize on the latest measurements for calculating the current airtime generated by the target coordinated devices:

$$\mathcal{C}oord_{(air)}(t_0) = \sum_{k=1}^{N_{CoordTerms}} (air_{k_{(\%)}}) \tag{10}$$

and subtract it from $\overline{\mathcal{DC}}_{\mathcal{F}_c}(t_0)$. The subtraction result is compared with the predicted class mean value of the current minute, as extracted from the ML long-term model. If the difference is greater than 10%, we consider that there is an active outlier, besides the coordinated terminals, which largely affects the $\mathcal{F}_c$. In such a case, the short-term $\hat{\mathcal{C}}_{max}$ calculation is enabled. In any other case, the proposed framework continues to calculate the $\mathcal{C}_{min}$, based on the ML prediction model described in Section III-B. Additionally, the latest values of the wireless analytics database participate in the calculation for $\mathcal{N}_{ContTerms}(t_0)$ and $\mathcal{C}ont_{(air)}(t_0)$. Alongside with $\mathcal{C}oord_{(air)}(t_0)$ and under the use of (8), the next minute's short-term $\hat{\mathcal{C}}_{max}$ prediction is calculated. This procedure is executed every minute and effectively protects the proposed system from short-term outliers as proved in Section IV.

IV. Evaluation

In this section, the evaluation of the proposed solution and the analysis of the collected results are described. It is worth to be noted that our framework is designed, developed and evaluated by exclusively using commercial off-the-shelf (COTS) hardware in a real testbed environment, under controlled and

uncontrolled external interference settings. More specifically, the indoor and outdoor testbeds of NITOS, located in the premises of University of Thessaly in Greece are used [44]. In this manner, we are able to prove the functionality and benefits of the framework, in both interference controlled (indoor) and fully uncontrolled (outdoor) spectrum environments. In the context of this work, we evaluate the proposed framework for different channel widths (20-40 MHz) in the 2.4GHz Wi-Fi band. In this band, given the country restrictions for the operation of devices in the 5GHz band, the phenomenon of over-congestion and interference presence is more intense, contrary to the 5GHz band. More specifically, the overlapping channels in parallel with the signal interference created by non-IEEE 802.11 devices, highly increase the $\mathcal{I}nt_{(Ext)}$ factor. In contrary, such a case is very unusual in the 5GHz frequencies due to the orthogonal channel allocation within the band. Of course, the proposed solution is applicable for either Wi-Fi bands and with channel widths 20-160 MHz. However, we opt to evaluate our framework under the most dense settings, in a disadvantageous and hostile spectrum environment. The results are organized as follows: in Section IV-A we present the actual throughput and $\mathcal{DC}$ utilization, which may be achieved for different modes of operation of IEEE 802.11 in real systems. In Section IV-B, we examine the accuracy of our solution, for the short-term measurements, utilizing the collected results from the previous subsection. Finally, in Section IV-C, we evaluate the long-term prediction framework, using either ML, or using ML and our self-correcting mechanism based on the outlier detection mechanism.

A. Real-World Considerations

Initially, the maximum theoretical throughput ($\mathcal{T}_{\mathcal{MCS}}$) is considered (eq: 2, 8), for predicting the estimated link's capacity. In the context of this work, we evaluate the proposed system under both TCP and UDP traffic requests. Thus, it is useful to compare how the experimentally achieved throughput ($\mathcal{A}_{thr}$), converges to the theoretical for each available MCS. More specifically, the proposed framework calculates the estimated capacity $\mathcal{C}_{min}$ and $\mathcal{C}_{max}$, as reported by a traffic generator tool (*iperf* [45]). In this manner, the estimated throughput reflects the results based on the goodput received at the application layer. Of course, lower values than the theoretical limits are expected, as the throughput at this layer does not contain overheads like IEEE 802.11 and IP headers, which are removed from lower OSI layers. Furthermore, it is very important to execute the experiments in completely isolated and unutilized frequencies. The deployed topology of this scenario corresponds to "Indoor Topology 1" in Fig. 1. Specifically, we deploy a strong signal link (-23dbm), between a Wi-Fi DU and a station, which are equipped with at least one of the Qualcomm Atheros (QCA) 93xx/986x/988x 802.11a/b/g/n/ac wireless network adapters. We carefully select the deployed links to perform as optimal as possible. All the examined links continuously achieved greater than 99.4% probability of successful transmissions during the experiments, at the highest MCS 23 for both 20 and 40 MHz channels. We configure the Wi-Fi DU to operate at primary channel 6 with HT40- capabilities and long Guard Intervals (GI). Serially, we configure all the available MCS profiles for the IEEE 802.11n standard and measure for 60 seconds the performance of each link, by creating saturated downlink TCP and UDP traffic streams with *iperf*. The averaged results after ten executions at each MCS, for both 20 and 40 MHz channel widths and 1-3 spatial streams (SS), are depicted in Fig. 6 and reported in Table I. Furthermore, two similar strong signal links with different wireless adapters were identically instantiated and tested, in order to validate the results obtained from the first deployed link. It is worth to be noted, that no significant deviations were observed between these three different examined links.

Additionally, the proposed model states the exploitable capacity of each channel to be "1". Practically, this means that an IEEE 802.11 device with no contention or external interference would be able to utilize 100% of the channel's airtime. However, based on the IEEE 802.11 standard, DCF Interframe Spaces (DIFS), Short Interframe Spaces (SIFS), Network Allocation Vectors (NAV), Contention Windows (CW) and other time depended parameters, some inevitable unexploited transmission slots exist in the Wi-Fi operation. In such way, it would be useful to experimentally measure these lost transmission opportunities and calculate the actual feasible channel airtime. During the calculation of $\mathcal{A}_{thr}$ as described in the previous scenario, we also examine the $\mathcal{DC}$ measurements for both primary and secondary operating channels per each MCS. In this manner, we observe the actually achieved channel utilization percentage to range from 90.91% to 96.17%. As this scenario involves the use of only two Wi-Fi devices, it is interesting to study if the participation of additional Wi-Fi DUs - STAs at the architecture, affects the aforementioned results. For this reason, we deploy additional links either at a random partially/fully overlapping channel with the primary one, which also transmit saturated traffic from the Wi-Fi DUs to STA(s). By altering the number of deployed Wi-Fi DUs from 1-5 and the associated STAs per DU in the range of 1-4, we observe that the $\mathcal{DC}$ values at the primary channel fluctuate from 90.58% to 96.70%. Finally, after averaging all the scenarios mentioned above, we calculate the total feasible channel utilization ($\mathcal{A}_{cc}$) to be 95.65%. Thereafter, we replace "1" with $\mathcal{A}_{cc}$ and $\mathcal{T}_{\mathcal{MCS}}$ with $\mathcal{A}_{thr}$ at the proposed model for remaining of our experiments.

B. Estimated Throughput Model Evaluation

In this experimental scenario, the proper function for the model proposed in Section III-C is extensively evaluated. More specifically, we assess the accuracy of the estimated throughput calculation in a target operating Wi-Fi channel. Likewise with the previous subsection, we obtain and evaluate the experimental results under the presence of controlled interference. In this context, we deploy a testbed node to act as the Monitor DU_2 and report the $\mathcal{DC}$ utilization values of the target channel, as well as the additional wireless analytics in real-time. These values alongside with those calculated in Section IV-A, are used to feed the proposed model and extract the estimated throughput

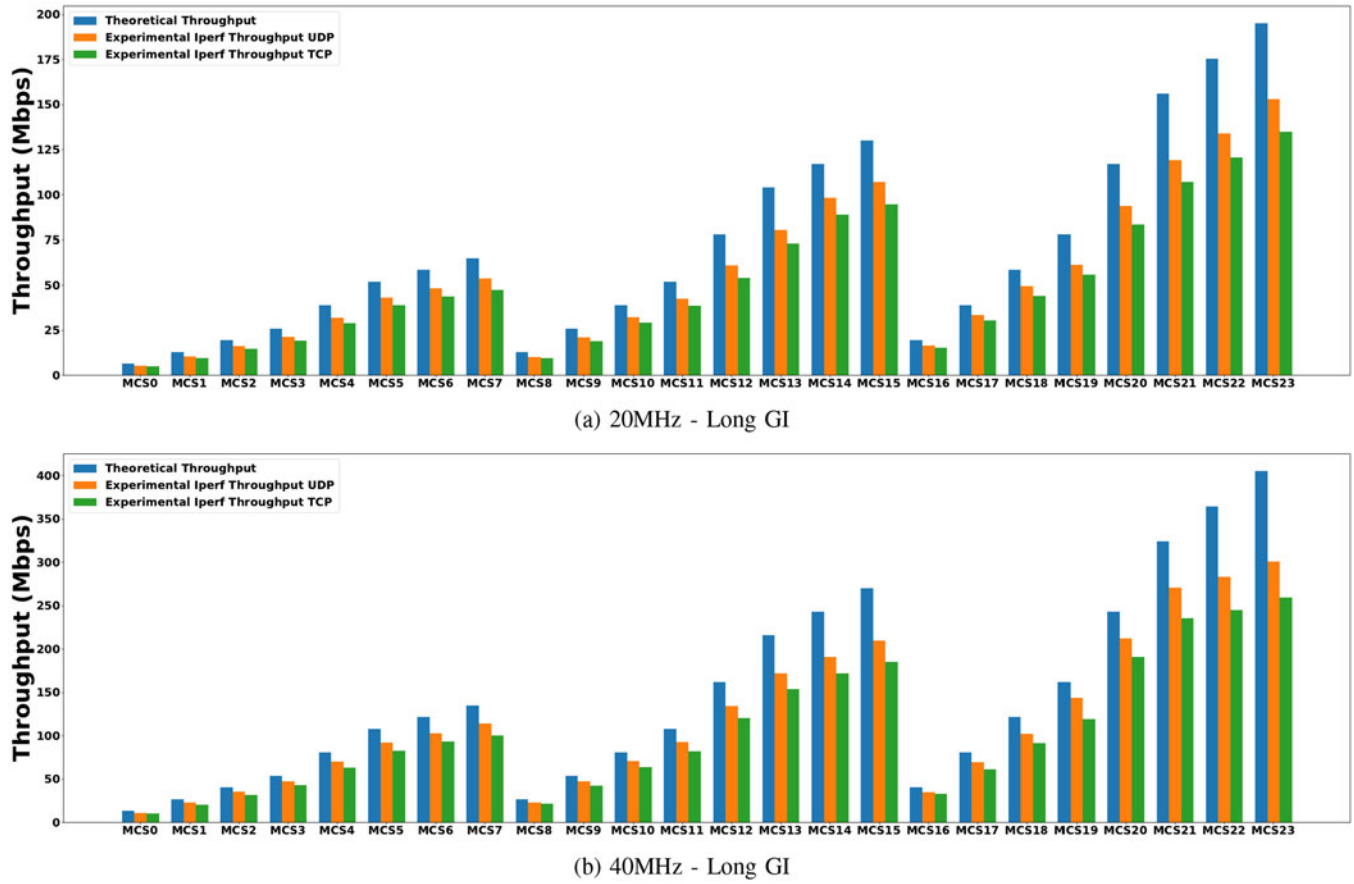


Fig. 6. (Indoor Topology 1) - Application Layer Goodput.

as proposed at Section III-C. The deployed under consideration "Indoor Topology 2," is analytically depicted in Fig. 1. For all illustrated cases here under, we configure the devices to use channel 6 (2437 MHz) with 20 MHz width acting as the the target operating $\mathcal{F}_c$ for the Wi-Fi DU_4 (**S**ystem **U**nder **T**est - SUT). It is worth to be noted that we select the aforementioned channel intentionally, as it can be considered the most vulnerable one, by receiving interference from both two sides of the channel, i.e. channels 2-5 and 7-10.

1) Non-Overlapping Interference: In this case, the interference resulting between two seemingly non-overlapping channels in the 2.4GHz band is analyzed and experimentally evaluated. Initially, we deploy the external interference link (Ext. AP_3) to operate in channel 1 and create medium UDP traffic of 25 Mbps using MCS15. As the theoretical operating frequencies of channels 1 (2402-2422 MHz) and 6 (2427-2447 MHz) do not overlap, no interference impact is expected in the under test link. However, Monitor DU_2 reports the $\mathcal{DC}_{\mathcal{F}_c}$ to be 14.1% for channel 6, which is considered directly as $\mathcal{I}nt_{(Ext)}$ with no contention opportunities. Based on the proposed model, the $\mathcal{C}_{max}$ is calculated to be 109.2 Mbps for MCS 22 used from Wi-Fi DU_4. In fact, the under-test link achieves 107.1 Mbps as depicted in Fig. 7(a), which is actually 15.7% lower than the $\mathcal{A}_{cc}$, as calculated in Section IV-A. For this scenario, the existence of interference as well as its impact, even for theoretically non-overlapping channels is highlighted. For this reason, the applicability of the proposed framework can also be considered essential for the 5GHz Wi-Fi band, in which the channels are orthogonal and therefore non-overlapping.

2) Saturated External Interference: In this scenario, we examine how accurate the proposed model can be, under the existence of heavy $\mathcal{I}nt_{(Ext)}$. More specifically, we set Ext. AP_2 and Ext. AP_3 to operate at channels 8 and 3 accordingly. Furthermore, saturated UDP traffic is generated from Ext. APs to their connected stations. The Monitor DU_2 reports the $\mathcal{DC}_{\mathcal{F}_c}$ to be 90.6% and $\mathcal{C}_{max}$ is calculated to be 6.76 Mbps. Veritably, the average throughput achieved at Wi-Fi DU_4 link is 6.68 Mbps and the experimental results are analytically shown in Fig. 7(b). It is obvious that the Ext. APs deplete any transmission opportunities for the Wi-Fi DU_4, by generating a high amount of interference at the target $\mathcal{F}_c$. As the proposed framework can distinctly separate $\mathcal{I}nt_{(Ext)}$ from $\mathcal{I}nt_{(Wi-Fi)}$, the actual transmission opportunities at each channel can be retrieved. Thus, based on the same experimental setup, the $\mathcal{C}_{max}$ is calculated to be 51.2 Mbps for channel 8 and 53.8 Mbps for channel 3. It is evident that if the proposed framework was utilized for channel selection schemes, it would lead to significantly higher performance for the SUT link.

3) Saturated Contention: For this case, the accuracy of $\mathcal{C}_{max}$ calculation is examined, when high contention exists in the same channel. More specifically, we deploy Ext. AP_1 to transmit saturated traffic in channel 6, towards generating

TABLE I
(INDOOR TOPOLOGY 1) - APPLICATION LAYER GOODPUT (MBPS)

MCS	20MHz - Long GI			40MHz - Long GI		
	Theoretical	*Experimental UDP*	*Experimental TCP*	*Theoretical*	*Experimental UDP*	*Experimental TCP*
0	6.5	5.23	4.93	13.5	11.3	10.6
1	13.0	10.6	9.52	27.0	23.2	20.8
2	19.5	16.2	14.6	40.5	35.3	31.5
3	26.0	21.3	19.2	54.0	47.4	43.1
4	39.0	32.1	28.8	81.0	70.2	63.1
5	52.0	43.1	38.8	108.0	92.4	83.0
6	58.5	48.2	43.7	121.5	103.0	93.7
7	65.0	53.8	47.5	135.0	114.0	100.5
8	13.0	10.2	9.62	27.0	23.1	21.7
9	26.0	21.2	19.0	54.0	47.4	42.6
10	39.0	32.4	29.1	81.0	70.8	63.8
11	52.0	42.7	38.7	108.0	92.6	82.4
12	78.0	61.0	54.2	162.0	134.0	120.6
13	104.0	80.4	73.1	216.0	172.0	153.5
14	117.0	98.2	88.8	243.0	191.0	171.9
15	130.0	107.0	94.5	270.0	210.0	185.0
16	19.5	16.4	15.3	40.5	34.9	32.8
17	39.0	33.6	30.6	81.0	69.3	61.6
18	58.5	49.6	44.1	121.5	102.0	91.7
19	78.0	61.4	55.8	162.0	144.0	119.5
20	117.0	93.8	83.5	243.0	212.0	190.8
21	156.0	119.0	107.1	324.0	271.0	235.7
22	175.5	134.0	120.6	364.5	283.0	244.7
23	195.0	153.0	134.9	405.0	301.0	259.1

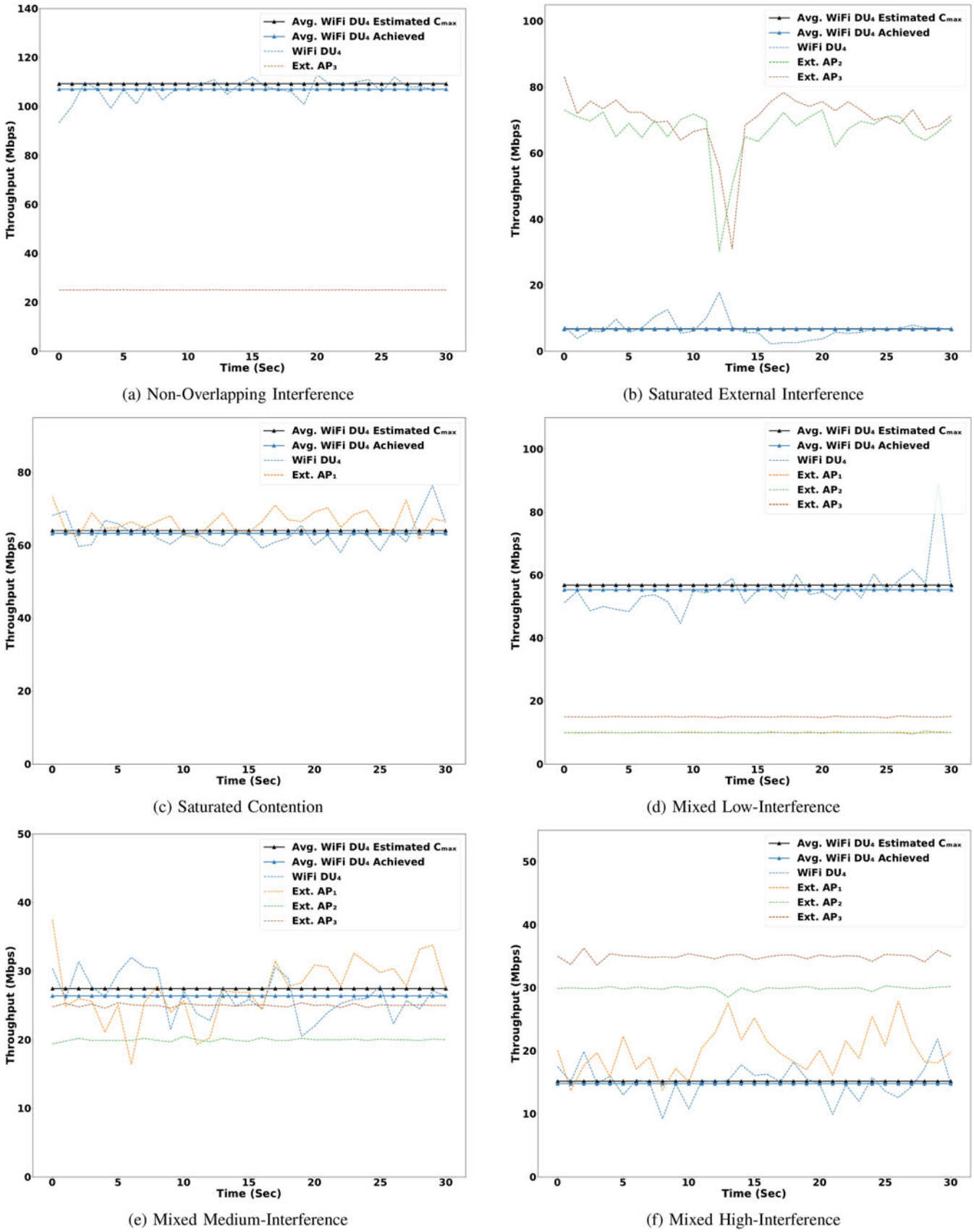
channel contention with Wi-Fi DU_4. The $\mathcal{C}_{max}$ is calculated to 64.07 Mbps for the Wi-Fi DU_4 while in practice the link achieves 63.3 Mbps. Fig. 7(c) illustrates that the two links share equally the medium as expected in the scenario of IEEE 802.11 and CSMA/CA contention. Almost identical performance can be observed, 63.3 Mbps for the Wi-Fi DU_4 and 66.5 Mbps for the Ext. AP_1, while both links operate using MCS 22.

4) Mixed Interference: Finally, three different scenarios with combined $\mathcal{I}nt_{(Ext)}$ and $\mathcal{I}nt_{(Wi-Fi)}$ are presented, for concluding the model's evaluation. Similarly with the previous, we keep Ext. AP_1 operating in channel 6 in order to create $\mathcal{I}nt_{(Wi-Fi)}$ to the SUT. Additionally, the Ext. AP_2 operates in channel 8 and the Ext. AP_3 in channel 3, thus generating various amounts of $\mathcal{I}nt_{(Ext)}$ to the Wi-Fi DU_4. As illustrated in Figs. 7(d), 7(e) and 7(f), three scenarios with low, medium and high amount of mixed interference are demonstrated.

Initially, we configure the Ext. AP_1 to transmit 10 Mbps of UDP traffic, in order to create $\mathcal{I}nt_{(Wi-Fi)}$ to the SUT's $\mathcal{F}_c$. Moreover, the Ext. AP_2 and Ext. AP_3 create 10 and 15 Mbps of UDP traffic respectively. In this manner, 53.2% of $\mathcal{DC}_{\mathcal{F}_c}$ is reported by the Monitor DU_2 with $\mathcal{C}_{max}$ estimated to be 56.88 Mbps. It is worth to be noted that, even though there is presence of $\mathcal{I}nt_{(Wi-Fi)}$ in this case, Ext. AP_1 can not be included in $\mathcal{N}_{ContTerms}$, since its total airtime (13.7%) does not conform with equation (6). Therefore, the achieved performance of Ext. AP_1 will not be affected from the saturated transmissions of the under-test link, which is indeed depicted in Fig. 7(d). In practice, the SUT achieved 55.3 Mbps during the experiment.

Following this, we increase the traffic demands for Ext. AP_1, Ext. AP_2 and Ext. AP_3 to 45, 20 and 25 Mbps accordingly. In this case, the Monitor DU_2 calculates the $\mathcal{DC}_{\mathcal{F}_c}$ to be 78.8% and the increased demands of Ext. AP_1 characterize now this terminal as $\mathcal{N}_{ContTerms}$. Thus, the SUT and Ext. AP_1 will equally compete for the available transmission opportunities in channel 6, with respect to the interference $\mathcal{I}nt_{(Ext)}$ caused by Ext. AP_2 and Ext. AP_3. This is also evident in Fig. 7(e), where the under-test link is shown to achieve achieve 26.4 Mbps, while the Ext. AP_1 gets 27.1 Mbps. Based on the proposed model, the $\mathcal{C}_{max}$ for the SUT is calculated to 27.47 Mbps.

Finally, we further increase the demands of Ext. APs to 55, 30 and 35 Mbps. Similarly with the previous case, the under-test link will contend for channel 6, while lower performance



(a) Non-Overlapping Interference

(b) Saturated External Interference

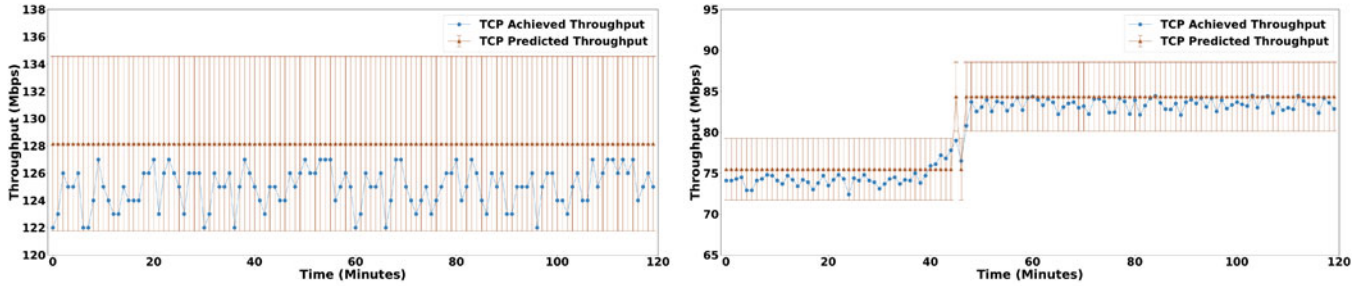
(c) Saturated Contention

(d) Mixed Low-Interference

(e) Mixed Medium-Interference

(f) Mixed High-Interference

Fig. 7. (Indoor Topology 2) - Estimated Throughput Evaluation.



(a) 06-Nov. 00:00 - 01:59 am (b) 06-Nov. 13:00 - 14:59 pm

Fig. 8. (Outdoor Topology 1) Long-Term Throughput Predictions.

is anticipated based on the increased demands. Fig. 7(f) illustrates that the experiment's results validate the expectations and the under-test link achieves 14.8 Mbps compared with the 15.17 calculated from the proposed model. Additionally, during this case, the competitive Ext. AP_1 achieved 19.2 Mbps.

In this subsection, several testbed scenarios were executed, in order to prove the proper functionality of the $\mathcal{C}_{max}$ throughput estimation model, introduced in Section III-C. Specifically, the estimation error varies from 1.19% to 3.9% across the scenarios, with a mean value of 2.24%. Therefore, it is verified that the proposed model is able to make accurate throughput estimations, under several challenging interference scenarios, which include the combination of $\mathcal{I}nt_{(Ext)}$ (from both overlapping and non channels) and $\mathcal{I}nt_{(Wi-Fi)}$.

C. Performance Evaluation Outdoor Topology

The following subsection aims to highlight the proper joint operation, of the performance estimation and prediction models, under entirely uncontrolled spectrum conditions. In this context, we design, develop and evaluate the examined scenario in the NITOS Outdoor Testbed, as shown in ("Outdoor Scenario 1") Fig. 1. Two stations were employed (UE_5 and UE_6), which are in close proximity with the Wi-Fi DU_5, thus achieving high rate transmissions. Specifically, UE_5 achieves MCS23 with the use of 3 antennas, while UE_6 MCS14 with 2 antennas. Of course, the number and placement of the depicted Ext. APs / Ext. UEs are completely unknown for the specific scenario and it is provided for the completeness of representation. Furthermore, the depicted Monitor DU_3 refers to the hardware which generates the spectral datasets, as described at the previous Sections.

1) General Observations: Initially, it is worth to state some observations, in order to fully comprehend the challenging spectral conditions given in the examined area. During the collection of the spectral measurements ($\mathcal{DC}$ + wireless analytics) at each execution, the Monitor DU_3 senses on average 114 present uncoordinated APs. 49 of these APs are categorized as potentially harmful, since they exceed the CCA threshold. Furthermore, on average 27 STAs are associated in the vicinity, 12 of which may potentially be active interferers. The above statistics describe the entire number of Wi-Fi terminals given at all available channels (1-11) of the 2.4GHz Wi-Fi band. It is also observed that 72% and 57% of the sensed APs and STAs, are operating in channels 1, 6 and 11. In such way, it is showcased that these three commonly used channels, can be largely affected from external interference. The Monitor DU_3 is deployed in the 3 rd floor of a building, in which the wider area is mainly composed of stores, offices and coffee shops. Thus, similar differences on the spectrum utilization were observed between weekdays and weekends, as well as between working and non hours. Finally, significant reduction at the $\mathcal{DC}$ utilization appeared at all channels, during the lockdown periods of the COVID-19 pandemic.

2) Long-Term $\mathcal{DC}$ Predictions Evaluation: The following experiments aim to showcase that the mechanism proposed in Section III-B is suitable for making accurate long-term utilization predictions, when severe short-term outliers are not present. More specifically, as long as the long-term predictions match the ongoing $\mathcal{DC}$ values, there is no need for activating the "Real-time Outlier Detection". Through the use of the RSSI value for each Wi-Fi DU / UE, which is available and updated every minute in the wireless analytics central database, we are able to calculate the $\mathcal{C}_{min}$ based on the long-term predicted $\mathcal{DC}$ values. In this case, the (2) is used, by replacing the $\mathcal{T}_{\mathcal{MCS}}$ with $\mathcal{A}_{thr}$, as described in Section IV-A and reported in Table I.

Initially, we consider the link Wi-Fi DU_5 - UE_5 and the time period between 00:00 - 01:59am for November 6th 2020. In order to get the reported values, saturated TCP traffic was generated during this experiment. The link was able to constantly achieve MCS23 and all the predicted $\mathcal{DC}$ values ranged between 2-9% for class 0-10. As described in Section III-B, we keep the mean $\mathcal{DC}$ value of the target predicted class, which in this case is equal to 5 for the entire duration of the experiment. Additionally, the use of the mean value can add an error at the $\mathcal{C}_{min}$ prediction, which corresponds up to $\pm 5\%$. In Fig. 8(a), it can be observed that the predicted values of the $\mathcal{C}_{min}$ fully match with the actually achieved throughput during the 2-hour execution of the experiment. Based on the mean predicted $\mathcal{DC}$ value, the $\mathcal{C}_{min}$ is calculated to be 128.15 Mbps. In practice, the link achieved in average 124.8 Mbps, with all values ranging within the admissible limits of $\mathcal{C}_{min}$ $\pm 5\%$ ($\pm$ 6.40 Mbps).

Thereafter, we also examine the Wi-Fi DU_5 - UE_6 link under saturated TCP traffic (MCS14) and between 13:00 - 14:59 pm of November 6th 2020. In this case, the deviation of the predicted $\mathcal{DC}$ values was between 3-16%, resulting in two different predicted classes (0-10) and (10-20). More specifically, from

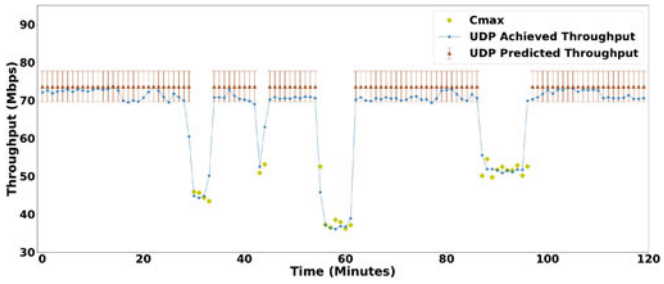


Fig. 9. (Outdoor Topology 1) Outlier Detection.

0 - 45 minutes, the model predicts a mean $\mathcal{DC}$ value 15 ($\mathcal{C}_{min} =$ 75.47 Mbps) and then reduces this utilization prediction to 5 ($\mathcal{C}_{min} =$ 84.36 Mbps). In Fig. 8(b), it is shown that during the two hour experiment only one missed detection from the proposed framework was identified. However, the deviation was not greater than "10%," which has been defined as the threshold for activating the "Real-time Outlier Detector," as discussed in Section III-D.

3) Outlier Detector Evaluation: Finally, the experimental results below showcase the proposed framework's operation under the presence of several short-term outliers. More specifically, the link Wi-Fi DU_5 - UE_6 is considered between 10:00 - 11:59 am, on December 23th 2020. The predicted values from the long-term ML model ranged between 23-29%, resulting in one predicted class (20-30). Thus the mean value of 25, was set for the entire experiment's duration. In order to evaluate the target link, saturated UDP traffic under the achieved MCS 14 was generated. Therefore, the predicted $\mathcal{C}_{min}$ is calculated to 73.65 Mbps with $\pm$ 3.68 Mbps error margin. The analytical results of the 2-hour experiment are summarized in Fig. 9. As it can be observed, 4 severe interferers of varied intensity and duration appeared during the experiment. For these cases, the proposed short-term mechanism was activated immediately and the $\hat{\mathcal{C}}_{max}$ values were extracted accurately. More specifically, the $\hat{\mathcal{C}}_{max}$ calculation mean error compared with the actual throughput achieved is 4.81%. As described in Section III-D, the current $\overline{\mathcal{DC}}_{\mathcal{F}_c}(t_0)$ is calculated through the $\mathcal{W}_{|n|}$ weight vector. An exponential weight vector could be applicable as well, by giving even greater emphasis on the most recent values. Hence, reduction of the $\hat{\mathcal{C}}_{max}$ mean error could be achieved. However, this could also result to short-term mechanism misuse when sharp peaks of $\mathcal{DC}$ appear. We consider the confrontation of this trade-off as future work.

D. Experimental Analysis

In this subsection, we provide a summary of the experimental findings illustrated in the previous subsections. Initially, Section IV-A describes some real-word considerations, which are calculated and taken into consideration, in order to properly calculate $\mathcal{C}_{min}$ and $\mathcal{C}_{max}$ values in real systems. Furthermore, Section IV-B includes an extensive indoor testbed evaluation for the proposed model. More specifically, Fig. 7 illustrates the results of several experimental scenarios and in which the high accuracy of $\mathcal{C}_{max}$ estimations is demonstrated. Several combined $\mathcal{I}nt_{(Ext)}$ and $\mathcal{I}nt_{(Wi-Fi)}$ interference scenarios were evaluated, in order to make sure that the proposed system, achieves high accuracy under various environments/conditions. Following, the purpose of Section IV-C is twofold. On the one hand, the accuracy of the proposed ML model is proved by accurately predicting the $\mathcal{C}_{min}$ values (Fig. 8). On the other hand, it is also highlighted that the mechanism proposed in Section III-D, is immediately activated when an short-term outlier is present and the $\hat{\mathcal{C}}_{max}$ values are precisely calculated (Fig. 9). It is worth to be noted, that all the outdoor scenarios are performed in a densely populated urban area where interference is completely uncontrolled.

V. Considerations

In this work, all the provided results under the examined experimental scenarios, represent the spectrum conditions from each Monitor DU's perspective. Concluding on the results, the proposed framework is able to export all the performance estimations and predictions. As the whole framework is based on Energy Detection methods, differences might occur between Monitor DUs and each Wi-Fi DU spectrum conditions. Of course, this is directly associated with the distance between the deployed Monitor DU and each Wi-Fi DU. In the examined scenarios, all the WiFi-DUs were in close proximity (1-7 m) with the Monitor DUs, thus eliminating the above phenomenon. However, the low cost and the plug & play installation can assist in the wide application of the proposed framework. More specifically, every Wi-Fi DU can be easily converted to Monitor DU by deploying an additional wireless adapter. The cost for the involved Atheros chipsets starts at $\sim$15$and can be installed at any conventional host machine through PCI or USB ports. In terms of software, the framework utilizes exclusively open source wireless drivers [46], [47] and open source software such as airodump, tcpdump and tshark. Therefore, the proposed framework can directly plug in any existing deployment. For denser and larger deployments, the proposed Monitor DU framework can be extended with more entities, each supervising a smaller region and providing accurate predictions in a common database environment. Such data can be used for forming radio interference maps, including predicted long-term values, allowing the configuration of each device in the most efficient manner. As the execution and the decisions made from the proposed mechanism are running on the operator side of the network, they do not require any software or hardware specific requirements on the STA side. Some interesting aspects which emerge in such architectures may contain the proper placement and channel selection for Wi-Fi / Monitor DUs, UE cell steering, as well as load balancing and social welfare maximization for devices operating in the unlicensed Wi-Fi bands. Recent research approaches can be found in literature, regarding channel selection and interference mitigation schemes [11], [12], [27], [48], which can be compared with the proposed work. Finally, greater emphasis at spectrum anomalies like capture effect and hidden terminals, can be given. There are already compatible mechanisms with our work, such as the work in [49], which aims in detecting such anomalies through the use of ML.

VI. Conclusion

In this work, a novel framework for estimating and predicting future interference levels at IEEE 802.11, was presented and experimentally evaluated. Analytical formulas for accurately calculating the available channel capacity $\mathcal{C}_{min}$ and $\mathcal{C}_{max}$ were developed and presented. By taking advantage of a ML approach, long-term $\mathcal{C}_{min}$ predictions can be extracted for the next 2-hours and with one minute accuracy. Additionally, a history-based short-term mechanism was developed, in order to shield our mechanism and when the ongoing spectrum conditions largely differ from the long-term predicted. These results are driven through the close monitoring of several wireless metrics, which are actively gathered from Monitor devices in our architecture and stored in central databases. Finally, extensive experiments at both Indoor and Outdoor network setups, showcased the accurate prediction and calculation for both $\mathcal{C}_{min}$ and $\mathcal{C}_{max}$, as well as the gains which can be obtained from the proposed work. In the future, we foresee to extend this work towards increasing the accuracy of the ML predictions for shorter-term periods, by considering changes in the ML model and predicting further metrics, stemming from the wireless analytics that we employ for channel capacity estimation. Moreover, tuning the scheduling algorithm of Wi-Fi, especially when using the Wireless Multimedia Extensions (WME), is expected to change the manner in which the short-term estimation is performed. Examining how this affects our decisions, and the overall framework lies within our future plans.

Acknowledgment

The European Union and its agencies are not liable or otherwise responsible for the contents of this document; its content reflects the view of its authors only.

References

[1] A. Pyattaev, K. Johnsson, S. Andreev, and Y. Koucheryavy, "3GPP LTE traffic offloading onto WiFi direct," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops*, 2013, pp. 135–140.

[2] 3rd Generation Partnership Project, "3GPP TS 36.360 v17.0.0 (2022-03), Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); LTE-WLAN Aggregation Adaptation Protocol (LWAAP) specification (Release 17)," Directory listing /FTP/specs/archive/36_series/36.360/, Mar. 2022. Accessed: Jun. 09, 2022. [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/36_series/36.360/

[3] 3rd Generation Partnership Project, "3GPP TS 23.402 V17.0.0 (2021-03), Technical Specification Group Services and System Aspects; Architecture enhancements for non-3GPP accesses (Release 17)," Directory listing /FTP/specs/archive/23_series/23.402/, Mar. 2021. Accessed: Jun. 09, 2022. [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/23_series/23.402/

[4] 3rd Generation Partnership Project, "3GPP TS 38.470 V17.0.0 (2022-04), Technical Specification Group Radio Access Network; NG-RAN; F1 general aspects and principles (Release 17)," Directory listing /FTP/specs/archive/38_series/38.470/, Apr. 2022. Accessed: Jun. 09, 2022. [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/38_series/38.470/

[5] N. Makris, C. Zarafetas, P. Basaras, T. Korakis, N. Nikaein, and L. Tassiulas, "Cloud-based convergence of heterogeneous RANs in 5G disaggregated architectures," in *Proc. IEEE Int. Conf. Commun.*, 2018, pp. 1–6.

[6] K. Chounos, S. Keranidis, T. Korakis, and L. Tassiulas, "Characterizing the impact of interference through spectral analysis on commercial 802.11 devices," in *Proc. IEEE Int. Conf. Commun.*, 2017, pp. 1–6.

[7] Y. Liu and J. Y. B. Lee, "An empirical study of throughput prediction in mobile data networks," in *Proc. IEEE Glob. Commun. Conf.*, 2015, pp. 1–6.

[8] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose, "Modeling TCP throughput: A simple model and its empirical validation," in *Proc. ACM SIGCOMM'98 Conf. Applications, technol., architectures, Protoc. Comput. Commun.*, 1998, pp. 303–314.

[9] Q. He, C. Dovrolis, and M. Ammar, "On the predictability of large transfer TCP throughput," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 35, no. 4, pp. 145–156, 2005.

[10] C. Wang, M. Di Renzo, S. Stanczak, S. Wang, and E. G. Larsson, "Artificial intelligence enabled wireless networking for 5G and beyond: Recent advances and future challenges," *IEEE Wireless Commun.*, vol. 27, no. 1, pp. 16–23, Feb. 2020.

[11] M. Peng, C. Kai, X. Cheng, and Q. F. Zhou, "Network planning based on interference alignment in density WLANs," *IEEE Access*, vol. 7, pp. 70525–70534, 2019.

[12] J. Mack, S. Gazor, A. Ghasemi, and J. Sydor, "Dynamic channel selection in cognitive radio WiFi networks: An experimental evaluation," in *IEEE Int. Conf. Commun. Workshops*, 2014, pp. 261–267.

[13] F. Maturi, F. Gringoli, and R. Lo Cigno, "A dynamic and autonomous channel selection strategy for interference avoidance in 80211," in *Proc. 13th Annu. Conf. Wireless On-demand Netw. Syst. Serv.*, 2017, pp. 1–8.

[14] P. Nayak, S. Pandey, and E. W. Knightly, "Virtual speed test: An AP tool for passive analysis of wireless LANs," in *IEEE INFOCOM 2019 - IEEE Conf. Comput. Commun.*, 2019, pp. 2305–2313.

[15] T. Y. Arif and R. F. Sari, "Throughput estimates for A-MPDU and block ACK schemes using HT-PHY layer," *J. Comput.*, vol. 9, no. 3, pp. 678–687, 2014.

[16] K. Mansour, I. Jabri, and T. Ezzedine, "Revisiting the IEEE 802.11n A-MPDU retransmission scheme," *IEEE Commun. Lett.*, vol. 23, no. 6, pp. 1097–1100, 2019.

[17] H. Lee, J. Cha, D. Kwon, M. Jeong, and I. Park, "Hosting AI/ML Workflows on O-RAN RIC Platform," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, 2020, pp. 1–6.

[18] T. Wang, S. Wang, and Z. H. Zhou, "Machine learning for 5G and beyond: From model-based to data-driven mobile wireless networks," *China Commun.*, vol. 16, no. 1, pp. 165–175, 2019.

[19] H. Fourati, R. Maaloul, and L. Chaari, "A survey of 5G network systems: Challenges and machine learning approaches," *Int. J. Mach. Learn. Cybern.*, vol. 12, no. 2, pp. 385–431, 2021.

[20] J. Moysen and L. Giupponi, "From 4G to 5G: Self-organized network management meets machine learning," *Comput. Commun.*, vol. 129, pp. 248–268, 2018.

[21] M. Polese, R. Jana, V. Kounev, K. Zhang, S. Deb, and M. Zorzi, "Machine learning at the edge: A data-driven architecture with applications to 5G cellular networks," *IEEE Trans. Mobile Comput.*, vol. 20, no. 12, pp. 3367–3382, Jun. 2020.

[22] YIN Rui, ZOU Zhiqun, WU Celimuge, YUAN Jiantao, CHEN Xianfu, and YU. Guanding, "Learning-based WiFi traffic load estimation in NR-U systems," *IEICE Trans. Fundamentals Electron., Commun. Comput. Sci.*, vol. 104, no. 2, pp. 542–549, Feb. 2021.

[23] N. Rastegardoost and B. Jabbari, "A machine learning algorithm for unlicensed LTE and WiFi spectrum sharing," in *IEEE Int. Symp. Dynamic Spectr. Access Netw.*, 2018, pp. 1–6.

[24] S. Kajita, H. Yamaguchi, T. Higashino, H. Urayama, M. Yamada, and M. Takai, "Throughput and delay estimator for 2.4GHz WiFi APs: A machine learning-based approach," in *Proc. 8th IFIP Wireless Mobile Netw. Conf. (WMNC)*, 2015, pp. 223–226.

[25] S. Chinchali and S. Tandon, "Deep Learning for Wireless Interference Segmentation and Prediction," 2012. Accessed: Jun. 09, 2022. [Online]. Available: http://cs229.stanford.edu/proj2012/ChinchaliTandon-DeepLearningforWirelessInterferenceSegmentationandPrediction.pdf

[26] J. Struye, B. Braem, S. Latré, and J. Marquez-Barja, "The CityLab testbed—Large-scale multi-technology wireless experimentation in a city environment: Neural network-based interference prediction in a smart city," in *Proc. IEEE INFOCOM 2018 - IEEE Conf. Comput. Commun. Workshops*, 2018, pp. 529–534.

[27] N. Makris, P. Karamichailidis, C. Zarafetas, and T. Korakis, "Spectrum coordination for disaggregated ultra dense heterogeneous 5G networks," in *Proc. Eur. Conf. Netw. Commun.*, 2019, pp. 512–517.

[28] "NITlab - Network Implementation Testbed Laboratory." Accessed: Jun. 09, 2022. [Online]. Available: https://nitlab.inf.uth.gr/

[29] K. Chounos, "Duty Cycle - monitor DU3 outdoor measurements (2021)," *IEEE Dataport*. Accessed: Dec. 20, 2021. [Online]. Available: https://dx.doi.org/10.21227/7cas-kv38

[30] "TP-LINK Archer C7 V2 Router," Accessed: Jun. 09, 2022. [Online]. Available: https://www.tp-link.com/us/support/download/archer-c7/v2/
[31] "IEEE 802.11ac MCS rates." Accessed: Jun. 09, 2022. [Online]. Available: https://community.cisco.com/t5/wireless-mobility-documents/802-11ac-mcs-rates/ta-p/3155920/
[32] H. Hewamalage, C. Bergmeir, and K. Bandara, "Recurrent neural networks for time series forecasting: Current status and future directions," *Int. J. Forecasting*, vol. 37, no. 1, pp. 388–427, 2021.
[33] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *Proc. 31st Youth Academic Annu. Conf. Chin. Assoc. Automat. (YAC)*, 2016, pp. 324–328.
[34] R. Dey and F. M. Salem, "Gate-variants of gated recurrent unit (GRU) neural networks," in *Proc. IEEE 60th Int. Midwest Symp. Circuits Syst.*, 2017, pp. 1597–1600.
[35] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," in *Proc. Ninth Int. Conf. Artif. Neural Netw. ICANN 99. (Conf. Publ. No 470)*, 1999, pp. 850–855.
[36] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Int. Conf. Learn. Representations*, 2014, *arXiv:1412.6980*.
[37] "Machine learning: Early stopping." Accessed: Jun. 09, 2022. [Online]. Available: https://keras.io/api/callbacks/early_stopping/
[38] *Airodump-ng*. Accessed: Jun. 09, 2022. [Online]. Available: https://www.aircrack-ng.org/doku.php?id=airodump-ng
[39] F. Fuentes and C. Dulal Kar, "Ethereal vs. Tcpdump: A comparative study on packet sniffing tools for educational purpose," *J. Comput. Sci. Colleges*, vol. 20, no. 4, pp. 169–176, 2005.
[40] *tshark - Dump and Analyze Network Traffic*. Accessed: Jun. 09, 2022. [Online]. Available: https://www.wireshark.org/docs/man-pages/tshark.html
[41] M. Bredel and M. Fidler, "Understanding fairness and its impact on quality of service in IEEE 802.11," in *Proc. IEEE INFOCOM 2009*, 2009, pp. 1098–1106.
[42] H.J. Escalante, "A comparison of outlier detection algorithms for machine learning," in *Proc. Int. Conf. Commun. Comput.*, 2005, pp. 228–237.
[43] R. Sparks, "Weighted moving averages: An efficient plan for monitoring specific location shifts," *Int. J. Prod. Res.*, vol. 42, no. 12, pp. 2521–2528, 2004.
[44] *NITOS - Network Implementation Testbed Using Open-Source Platforms*. Accessed: Jun. 09, 2022. [Online]. Available: http://nitos.inf.uth.gr/
[45] "iPerf - The Ultimate Speed Test Tool for TCP, UDP and SCTP." Accessed: Jun. 09, 2022. [Online]. Available: https://iperf.fr/
[46] *Atheros Linux Wireless Drivers - Ath9k*. Accessed: Jun. 09, 2022. [Online]. Available: https://wireless.wiki.kernel.org/en/users/drivers/ath9k/
[47] *Atheros Linux Wireless Drivers - Ath10k - Ath10k Backports Releases*. Accessed: Jun. 09, 2022. [Online]. Available: https://wireless.wiki.kernel.org/en/users/drivers/ath10k/backports/
[48] K. Chounos, N. Makris, and T. Korakis, "Enabling distributed spectral awareness for disaggregated 5G ultra-dense HetNets," in *Proc. IEEE 2nd 5G World Forum*, 2019, pp. 304–309.
[49] I. Syrigos, N. Sakellariou, S. Keranidis, and T. Korakis, "On the employment of machine learning techniques for troubleshooting WiFi networks," in *Proc. IEEE 16th Annu. Consum. Commun. Netw. Conf.*, 2019, pp. 1–6.

Kostas Chounos received the bachelor's degree in applied informatics engineering from the Technological Educational Institute of Crete, Heraklion, Greece, in 2010, and the master's and Ph.D. degrees in science and technology of computers and telecommunications engineering from the Department of Electrical and Computer Engineering (ECE), University of Thessaly, Volos, Greece, in 2014 and 2021, respectively. He is currently a Postdoctoral Researcher with the Department of ECE, University of Thessaly. Since 2013, he has been participating in several national and EU-funded research projects with the University of Thessaly. His research interests include wireless networks, cognitive radio networks, interference mitigation, and wireless resource allocation.

Panagiotis Karamichailidis received the B.Eng. degree in computer, communications, and network engineering, and the M.Eng. degree in electrical and computer engineering from the University of Thessaly, Volos, Greece, in 2018 and 2020 respectively. He is currently working toward the Ph.D. degree in computer science with the King Abdullah University of Science and Technology, Thuwal, Saudi Arabia. His research interests include cybersecurity, Internet of Things and cyber-physical security, network security, critical infrastructure-industrial control systems security and Machine Learning for networks/communications.

Nikos Makris received the bachelor's degree, the master's degree in computer science and communications, and the Ph.D. degree in electrical and computer engineering from the Department of Electrical and Computer Engineering, University of Thessaly, Volos, Greece, in 2011, 2013, and 2020, respectively. He is currently a Postdoctoral Researcher with the Centre for Research and Technology Hellas (CERTH) and the Department of Electrical and Computer Engineering, University of Thessaly. Since 2011, he has been participating in several EU-funded collaborative research projects with CERTH and University of Thessaly. His research interests include experimentally driven research with several radio access technologies (WiFi, WiMAX, LTE, 5G-NR), conducted under real environment settings, the disaggregation of the telecommunications network and subsequent management of the network using machine learning, multi-access edge computing, and NFV orchestration using open source platforms.

Thanasis Korakis received the B.S. and M.S. degrees in computer science from the University of Athens, Zografou, Greece, in 1994 and 1997, respectively, and the Ph.D. degree in electrical engineering from the University of Thessaly, Volos, Greece, in 2005. In summer of 2004, he was a Visiting Researcher with the CSE Department, University of California, Riverside, CA, USA. From 2005 to 2006, he was a Research Scientist with ECE Department, Polytechnic University, New York, NY, USA, where he was a Research Assistant Professor from 2006 to 2012. He is currently an associate professor with the ECE Department, University of Thessaly. His research interests include networking field with an emphasis on access layer protocols, cooperative networks, quality-of-service provisioning, network management, and experimental platforms. From 2007 to 2012, he was a Voting Member of the IEEE 802.16 Standardization Group. He was the recipient of several awards, including the Best Paper Awards in WiNTECH 2013, GREE 2013, and CloudComp 2015.