D1.2 WP1 Research Report I

Project Name: Anticipatory Networking Techniques in 5G and Beyond
Acronym: ACT5G
Project no.: 643002

Start date of project: 01/05/2015 **Duration:** 48 Months

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Actions.

Document Properties

Document ID	EU-H2020-MSCA-ITN-2014-643002-ACT5G-D1.2
Document Title	D1.2 WP1 Research Report I
Contractual date of delivery to REA	Month 24
Lead Beneficiary	Linköpings universitet(LiU)
Editor(s)	D. Yuan – LiU
Work Package No.	1
Work Package Title	Network Anticipation
Nature	Report
Number of Pages	7
Dissemination Level	**PUBLIC**
Contributors	LiU: V. Angelakis, D. Yuan POLIMI: A. Capone, M. Cesana Bell labs: I. Malanchini
Version Number	1

Contents

0 Executive Summary

This report details the work progress within work package one Network Anticipation of the ACT5G project. More specifically, the document provides information of the conducted and expected work of early-stage researcher (ESR) one and two. The document first gives an overview of the focus area and research topics. Technical details of the work are then presented by means of research paper to be published in 2017.

1 Work Plan and Progress of ESR 1

The research activities of the ESR 1 include the development of prediction and anticipation algorithms based on the analysis of information coming from real mobile radio networks, including data originating form network counters, probes and drive tests. The main goals of the ESR are twofold: first, to design and apply machine learning tools which can be exploited to derive compact and insightful Key Performance Indicators (KPIs), second to leverage the knowledge extracted from real data to anticipate and predict the behaviour of the mobile radio networks in order to optimize their performances

In this context, the ESR 1 Daniel Martin Weibel started working in two different directions: first, to get acquainted with quantitative tools of machine learning, he participated in a research activity to exploit real data traces coming IEEE 802.11-based campus networks to infer the habits and behaviour of network users; working on this research line Daniel Martin Weibel acquired skills in manipulating network-based data and using supervised and unsupervised machine learning tools. The research activity generated a publication which is included in the Appendix.

In parallel, Daniel Martin Weibel work to narrow down a research line of data-driven resource allocation in mmWave mobile radio networks, with the main goal to design resource allocation algorithms by leveraging large data sets of physical layer measures to be collected in real mmWave testbeds.

While working in this direction Daniel Martin Weibel notified the project management board about his intention to quit the project and the related PhD programme at PoliMI. Daniel Martin Weibel then formally quit his collaboration with the project on November, 1 2016.

A new recruiting process was then launched, leading to the recruiting of a new Early Stage Researcher 1, Claudia Parera Sotolongo. The recruiting process closed in February 2017 and the new ESR 1 will officially join the PhD program in Information Technology at the beneficiary Politecnico di Milano on May 1, 2017. Claudia is expected to follow the direction of the use of machine learning for the analysis of large quantity of quality-indicating data from mobile networks for anticipation.

2 Work Plan and Progress of ESR 2

The research objective of ESR 2 consists in the development of models and algorithms that are able to estimate and predict network performance for 5G systems. To this end, the work investigates the prediction of network coverage, rate, and other performance indicators at various time scales, in order to improve the operation quality and efficiency with an anticipatory-enabled framework. A number of tools, such as stochastic geometry and compressive sensing, are of interest to be considered for the purpose of prediction. Thus the ESR is expected to acquire fundamental knowledge of such tools, as well as to apply them to specific network scenarios. The study shall address the timescale, accuracy, and complexity of the tools for the scenarios in question.

Having the above in mind, ESR 2, Cristian Tatino, has first performed a survey of the literature on estimation and prediction. The next step has been the identification of a 5G system setup to focus the study on. After several discussions with the supervisors, Cristian has chosen millimetre wave (mmWave) system for the first study of prediction techniques. The type of systems is part of the 5G evolution, by means of transmissions using frequency in the range of 30-300 GHz. As there is a large amount of spectrum resource available, the data rate can reach very high levels. On the other hand, at such high frequencies, large path loss and link blockage effects pose major challenges. Thus models predicting the performance of mmWave are highly development for understanding the potential (and possible limitations) of mmWave.

Based on the knowledge acquired stochastic geometry, Cristian Tatino has developed a novel stochastic model for characterizing the coverage probability of a beam for outdoor mmWave systems. A feature of this analytic model is the consideration of reflection. In mmWare systems, the coverage is mainly provided by the beam directing to the user. However, reflections can be exploited in non-line-of-sight conditions to mitigate the blocking effect of obstacles lying on the direct path between the base station and the user. Comparing the numerical results obtained by the analytical model with simulations, it is concluded that the proposed model is able to capture of relation between coverage and the distribution and density of obstacles as well as the width of the beam used. The work has resulted in a conference paper, accepted for presentation and publication by the Spatial Stochastic Models for Wireless Networks (SpaSWiN) workshop of the 15th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT) that will take place in May, 2017.

As next steps, Cristian Tatino is 1) extending the model for additional scenarios, such as multi-connectivity, and 2) considering beam design consisting of the beamwidth, direction, and power per beam, along with allocation of the beams to users taking into account the anticipatory aspect.

3 Appendices:

- Understanding the WiFi Usage of University Students
- Beam Based Stochastic Model of the Coverage Probability in 5G Millimeter Wave Systems

Understanding the WiFi usage of university students

Alessandro E. Redondi, Matteo Cesana, Daniel M. Weibel
Dipartimento di Elettronica, Informazione e Bioingegneria
Politecnico di Milano
Milan, Italy
Email: {name.surname}@polimi.it Emma Fitzgerald
Department of Electrical and Information Technology
Lund University
SE-221 00 Lund, Sweden
Email: emma.fitzgerald@eit.lth.se

Abstract

In this work, we analyze the use of a WiFi network deployed in a large-scale technical university. To this extent, we leverage three weeks of WiFi traffic data logs and characterize the spatio-temporal correlation of the traffic "signals" at different granularities (each individual access point, groups of access points, entire network). The spatial correlation of traffic signals across nearby access points is also assessed. Then, we search for distinctive fingerprints left on the WiFi traffic by different situations/conditions; namely, we answer the following questions: *Do students attending a lecture use the wireless network in a different way than students not attending a lecture?*, and *Is there any difference in the usage of the wireless network during architecture or engineering classes?* A supervised learning approach based on Quadratic Discriminant Analysis (QDA) is used to classify empty vs. occupied rooms and engineering vs. architecture lectures using only WiFi traffic logs with promising results.

Index Terms

WiFi data analysis, user behaviour analysis,

I. Introduction

Wireless local area networks (WLANs) based on the IEEE 802.11 standard family (i.e., WiFi) are an essential building block to provide widespread wireless connectivity in diverse indoor/outdoor scenarios. The cities we live in, our workplaces, hospitals and other public and private buildings are equipped with WiFi networks to provide hot spot or capillary wireless connectivity.

The reasons for the success of WiFi range from the use of unlicensed spectrum to their ease of deployment and management [1]. Nowadays, WiFi connectivity is available both in users' terminals (laptops, handheld devices, electronic gadgets) as well as in embedded devices and appliances. To this extent, WiFi plays a key role in the provision of connectivity in urban environments and to fully realize the vision of smart cities.

Besides their primary role of providing connectivity, WiFi networks and devices nowadays come with powerful monitoring systems able to collect and store large quantities of data on the behaviour of the network itself: traffic load, number of users, quality of the wireless signal etc. Such data, which is primarily used for network management, optimization, and fine-tuning, is also a "goldmine" for offering byproduct services; indeed, WiFi logs can be used to provide location-based services by properly localizing the users and/or to estimate flows and spatial distributions of people during events or at shopping malls. Moreover, WiFi logs can be coupled with other types of context data and, more generally, can be used to assess the behaviour of users. Therefore, a careful analysis of such data provides valuable information that can be used for several purposes. In particular, in the context of smart cities and smart buildings, the availability of techniques for extracting high-level information from network traces may help city and building administrators to better understand and react to the citizens' needs.

In this work, we focus on a particular type of building present in many cities in the world, namely a university campus building. We analyze the data coming from the local WiFi network of Politecnico di Milano, a large-scale technical university located in Italy and we provide the following contributions: (i) we propose a temporal and spatial characterization of the WiFi traffic; (ii) we leverage the WiFi traffic traces to answer the following questions: *Do students attending a lecture use the wireless network in a different way than students not attending a lecture?*, and *Is there any difference in the usage of the wireless network during architecture or engineering*

classes? To answer these two questions we first propose a set of features which combine different attributes of the WiFi traffic (number of users, number of active connections, duration of active connections, etc.). Then we build up a labeled data set by exploiting information from the facility management department database which allows us to know if a room is occupied by a given lecture in a given time slot. Finally we propose a supervised learning approach based on Quadratic Discriminant Analysis (QDA) to classify empty vs. occupied rooms and engineering vs. architecture lectures by observing related WiFi traffic.

This work is organized as follows: Section II reviews the relevant literature on WiFi network data analysis; the reference scenario and spatio-temporal analysis of the WiFi traffic is presented in Section III. Section IV describes the supervised learning approach to classify empty vs occupied rooms and engineering vs architecture lectures. Section V concludes the manuscript and comments on ongoing/future work.

II. Related Work

The analysis of data traces extracted from WiFi networks has attracted much attention in the scientific community over the last decade. The relevant work in the field can be classified by the main target of the data analysis.

The first class focuses on general performance analysis and characterization of WiFi-based wireless networks with the goal of medium/long-term network optimization. In their seminal work [2], Kotz and Essien focus on the analysis of a WiFi campus network composed of 476 Access Points. The reference data set spans a two-month period and is composed of traffic- and association-related information collected through SNMP polling and SYS log messaging. The *active* data collection is further complemented by data collected by passive sniffers to capture back-end traffic. The collected data set is leveraged to perform a rather complete analysis on the traffic load characteristics (per user, per access point traffic, traffic variability over time, per building traffic), the traffic type characteristics (traffic breakdown per application) and user mobility (number of visited Access Points while associated). A similar analysis is performed on the same WiFi network in [3] after two years to assess the changes over time of the aforementioned performance figures. A performance analysis of a WiFi campus network is also targeted in [4].

Calabrese *et al.* use in [5] the wifi traces collected within the MIT WiFi network (3000

access points) to perform spatio-temporal analysis of the traffic flowing through the access points. Moreover, the information on the number of connected users per access point over time is used to classify location-dependent network behavior; namely, the authors show that by applying standard clustering techniques, location-dependent "fingerprints" can be determined for the network behavior (number of users, traffic).

The analysis of a corporate WiFi network is addressed in [6]; the authors collect a one-month data trace by polling every 5 minutes via SNMP 177 Access Points over three distinct corporate buildings. Besides analyzing the traffic characteristics, the authors also propose a clustering approach for the users based on two features: the *prevalence* which accounts for the time a user spends at an access point and the *persistence* which captures the total consecutive time a user spends at a AP.

The analysis of outdoor commercial WiFi networks is addressed in [7], [8] and [9]. In [7], Blinn*et al.* consider a Verizon WiFi Hot Spot network composed of 312 access points running IEEE 802.11b which are polled via SNMP every 5 minutes to obtain traffic and load measures. The proposed analysis is mainly targeted to characterize the access point traffic distribution over time, further assessing the spatial correlation existing among adjacent access points. A user mobility analysis is also carried out by leveraging the same clustering framework proposed in [6]. Afanasyev *et al.* [8], [9] address the performance analysis of the Google WiFi network in Mountain View (CA), composed of 500 access point. Different from the previous works, the reference wireless network architecture also has a wireless mesh extension interconnecting the access points. The proposed analysis has three main contributions: (i) the characterization of per-user traffic distributions, (ii) the classification of users in terms of their pattern in network usage and generated traffic (sporadic users, residential users, etc.), (iii) the assessment of users mobility in terms of travelled distance distributions.

Ganji *et al.* leverage in [10] campus WiFi data traces collected through SNMP polling and SYSlog messaging to evaluate the potential savings of duty cycling management policies of the access points in large-scale campus wireless networks.

A second class of work explicitly targets the assessment of human mobility and the development of data-driven mobility models [11]. Song *et al.* takes as reference the same data traces used in [2] to build up and validate location predictors for WiFi users of campus networks; two approaches are introduced based on Markovian models and Lempel-Ziv predictors. Along the

same lines, in [12] and subsequent studies [13], [14], Kim *et al.* introduce and evaluate mobility models for campus WiFi users which are developed and trained on WiFi data traces.

Mobility analysis in hospitals is the main focus in [15] and [16], where movements of people are estimated by leveraging the data coming from a WiFi network deployment composed of 798 access points at Aarhus hospital. Besides characterizing the users mobility, the authors also propose a clustering analysis to classify different types of users (medical staff, visitors, patients, etc.).

The third and last class of works include the literature where WiFi data traces are mainly leveraged to classify the type of users. In [17] four months of WiFi traces are collected in a campus environment and used to characterize the different behavior in terms of mobility between smartphones and laptops. Gember *et al.* extend the comparative analysis between handheld and non-handheld devices by characterizing the different users behaviors in terms of type of generated traffic; namely, full packet traces collected in a WiFi campus environment are collected and analyzed to assess the applications/traffic breakdown for handheld and non-handheld devices.

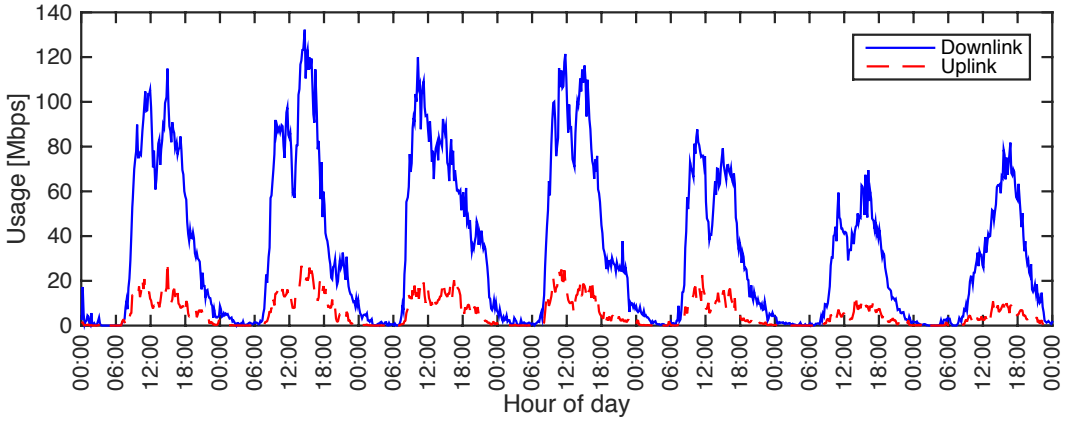
Recently, Wei *et al.* analyzed in [18] the impact of the high penetration of handheld devices in WiFi campus networks. A one year-long DHCP log is used along with a one-month flow level data trace to profile handheld and non-handheld devices in terms of mobility behavior and traffic characteristics.

According to this characterization, this paper stands in the third class.

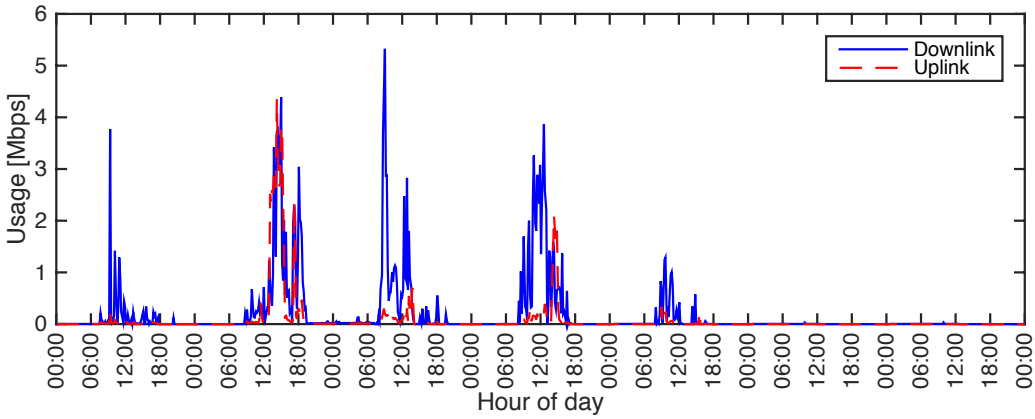
III. Scenario

This work analyses network data traces extracted from the wireless network of the architecture department building of the Politecnico di Milano (PoliMi) university, located in Milan, Italy. The wireless network under study is composed of 28 different access points (APs) located on four different floors of the building and covering rooms devoted to lectures as well as offices, corridors and other public spaces. The APs support the AirWave Management Platform (AMP), that allows to observe every device and user connected to the network. In particular, the AMP allows to sample the bandwidth usage and the number of clients connected for each APs every 5 minutes.

In this work, we focus on a period of three weeks, from the 16th of November, 2015 to the 6th of December, 2015. This period does not contain any national holiday, therefore it represents well the “steady-state” behaviour of the wireless network in terms of number of clients connected



(a)



(b)

Fig. 1. (a) Cumulative downlink/uplink usage over one week; (b) Downlink/uplink usage of one particular AP over the same week

and average usage. For each access point, the uplink/downlink bandwidth usage and the number of clients connected were sampled and stored every 5 minutes, for a total of $12 \times 24 \times 21 = 6048$ data points per AP. Additionally, the list of devices connected to each AP is downloaded and stored in a database. For each connected device, the following information is stored and updated every 5 minutes: device MAC address, timestamp of the association with the AP, duration of the connection, average and variance of the signal quality of the connection [dB] as well as average and variance of the bandwidth usage [kbps]. Over the three weeks, a total of 27538 unique devices were observed, generating 300681 different connection events.

Figure 1(a) shows the cumulative uplink/downlink bandwidth usage summed over all the APs during one of the selected three weeks. As one can see, such cumulative measures show a very nice periodic behaviour: the network usage is very low during the night, increases rapidly in the morning, experiences a short decline during lunch break and fades out during the afternoon with a short constant period of about two hours during the evening. This pattern is clearly associated with the behaviour of students and employees working in the building under consideration, and repeats unchanged every day from Monday to Friday. The behaviour during the weekend is similar to the one during work days, although with fewer clients connected and therefore less bandwidth usage. The same behaviour is obtained for the other two weeks in the dataset, which are not shown here for space reasons. Unfortunately, the nice periodic behaviour observed on the cumulative data is no more visible when analysing the traces of a particular AP. As an example, Figure 1(b) shows the downlink and uplink usage measured on an AP located in a room used for lectures throughout the week. As one can see, the bandwidth usage is different from day to day and no periodic pattern (except from the trivial night/day differences) may be found. Such consideration, that periodicity is lost when going from cumulative to individual AP data, is similar to what was found in [19] for cellular networks.

In order to provide a more detailed analysis of the behaviour of the network, we investigate how the network load varies both spatially and temporally.

A. Temporal correlation analysis

As a first step, we compute the sample Pearson correlation coefficient $r_i(t)$ of the downlink/uplink usage of the i-th AP with a version of itself delayed by t samples (the so called *lagged* correlation). This is done to determine if some temporal pattern still exists in individual AP data. Figure 2 shows the average lagged correlation coefficient for the downlink bandwidth usage, where the average is taken over all APs. As one can see, the average temporal correlation is low, with peaks occurring every 12 and 24 hours according to the diurnal human activity pattern. It is interesting to note that the correlation coefficient has local maxima in correspondence to lags 168 and 336, that is after one and two weeks. This can be explained taking into account the schedule of lectures in some of the rooms covered by the APs, which has a periodicity of one week. The 95% confidence interval (red dashed line) is very narrow, meaning that all APs show similar correlation coefficients. We repeated the same test for the uplink traffic (not shown here

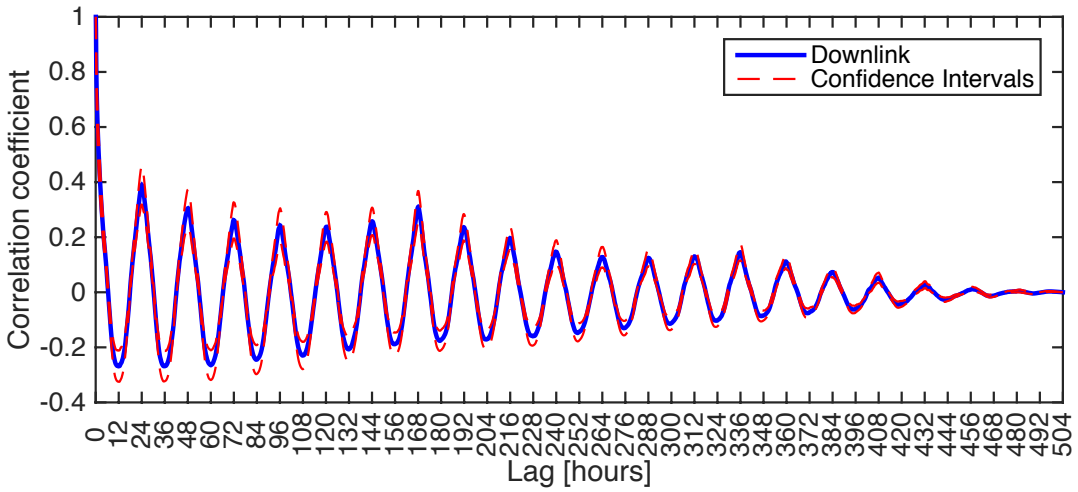


Fig. 2. Temporal correlation averaged over all APs. Local maxima are visible after 7 and 14 days, according to the academic calendar.

for space reasons), finding weaker temporal correlation although with the same periodic pattern.

B. Spatial correlation analysis

We also analyse the spatial correlation between different APs, that is, we compute the correlation coefficient $r_{i,j}$ for all possible pairs (i, j). Figure 3 shows the computed correlation coefficients for the downlink and uplink traffic, where blue indicates no correlation ($r_{i,j} = 0$) and yellow indicates maximum correlation ($r_{i,j} = 1$). The four red squares indicate correlation values of APs located on the same floor (basement, ground, first and second floor, from the top left) and the AP with index 13 is located alone on a mezzanine. Some interesting observations can be made from the inspection of Figure 3:

1) The maximum correlation is observed among APs on the first and second floor, which are also those where most of the classrooms are located. Overall, the maximum spatial correlation value for the downlink traffic is 0.7, while it is 0.5 for the uplink case (again indicating a weaker correlation in the uplink usage).
2) Conversely, APs located in the basement and on the ground floor, exhibit weak spatial correlation.
3) Some particular APs (rows 8, 11, 13 and 17) show very low correlation values with all other APs. A more thorough analysis revealed that these APs show the lowest uplink/downlink traffic usage and the lowest average signal quality towards connected devices. This may

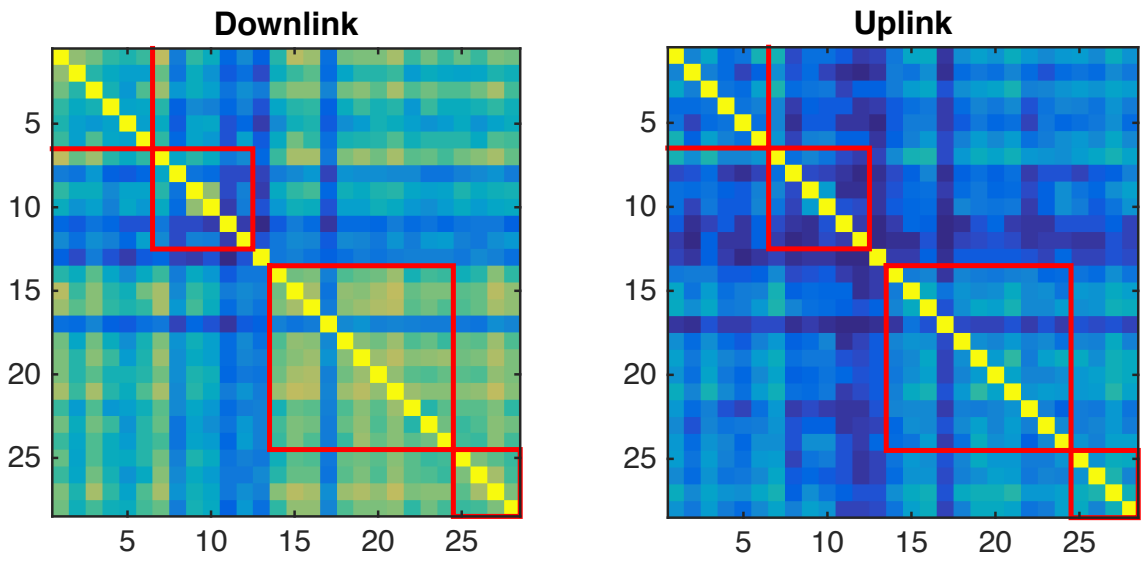


Fig. 3. Spatial correlation among different APs.

suggest that either they are malfunctioning or that their positioning may be changed to improve their performance.

IV. Classification

The wireless network under study is located in a building belonging to the Department of Architecture of PoliMi university. This building contains several classrooms that are used for teaching classes from both the Architecture and Engineering courses of study. Each classroom is different in size (maximum number of students it may host) and is used for lectures only in specific time slots during each day, according to the academic calendar. When a room does not host a lecture, it may be still used by students for studying or just for passing time between a lecture and the following one.

In such a scenario, we rely on the data available AirWave Management Platform and corresponding to those APs located in the classrooms to answer two different questions:

1) Is there a difference between the WiFI usage inside rooms during lecture times and "idle" times of these rooms?
2) Is there a difference between the WiFi usage inside rooms during architecture lectures and engineering lectures in these rooms?

The answer to the first question is somehow expected as positive, as it seems logical to assume that students attending a class will pay more attention to the teacher rather than to their smartphones or laptops. At the same time, a user studying or passing time in an empty room may be using its wireless connection actively. Conversely, the answer to the second question seems more unpredictable: although engineering and architecture students tend to consider themselves as different "species", the way they behave during classes in terms of network usage may be the same.

A. *Feature extraction*

To answer these questions, we take the following approach. First, we rely on the academic calendar to identify when and in which classroom architecture and engineering lectures are given. This allows to extract several "*time slots*" from the data traces of the APs located in the rooms where lectures are given, where each time slot corresponds to either a particular lecture (from the architecture or engineering courses) or to an empty slot (see Figure 4). Note that we do not consider empty slots occurring during night hours, but only those daytime periods in which a room is not used for a lecture. Therefore, we only consider time slots falling between 7 AM and 7 PM. Over the three weeks under study, we extracted 213 non-empty time slots (156 corresponding to architecture lectures and 57 to engineering lectures) and 101 empty time slots.

For each time slot, we extract the following features:

- *Total number of connections:* the total number of connections whose start time falls inside the time slot.
- *Number of unique devices:* the number of unique MAC addresses that started a connection in the time slot.
- *Number of connections per device:* the ratio between the total number of connections and the number of unique MAC addresses seen in the time slot, or how many times a particular device connected to the network in the time slot.
- *Average normalised duration:* the average duration of each connection, divided by the duration of the time slot.
- *Average and variance of bandwidth usage:* the mean and variance of the bandwidth usage of each connection, averaged over all the connections in the time slots. These features capture

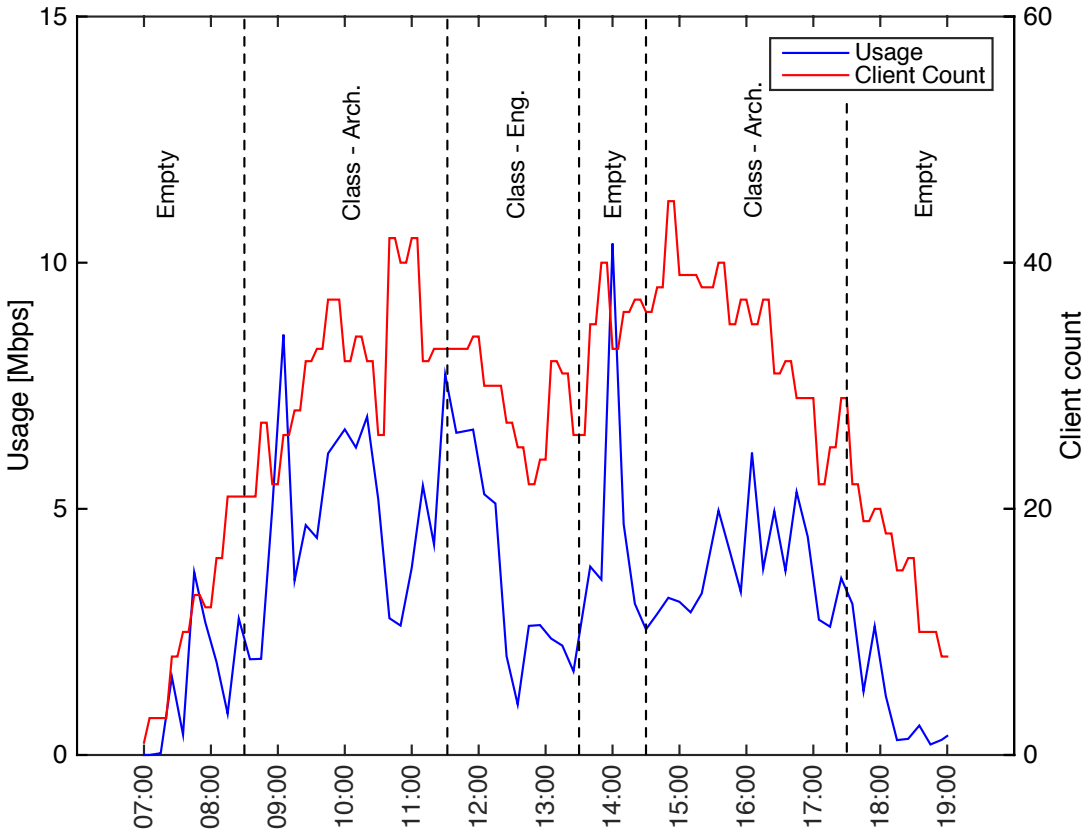


Fig. 4. Segmentation of network data into different time slots according to the academic calendar.

how much traffic is produced by students in a particular time slot and how variant this traffic is on average.

- *Average and variance of signal quality:* the mean and variance of the signal quality experienced by each connection, averaged over all the connections in the time slots.
- *Occupancy:* the ratio between the number of unique devices seen during the time slot divided by the seating capacity of the room in which the particular AP is installed.
- *Normalised occupancy:* the occupancy value normalised with respect to the duration of the time slot.
- *Connection distribution sparsity and peak:* for each time slot, we observe the particular minute at which each connection is started and we build an histogram with bins spaced every five minutes, for a total of 12 bins. We then compute the distribution sparsity and peak as the number of empty bins and as the index of the bin with the highest connection count,

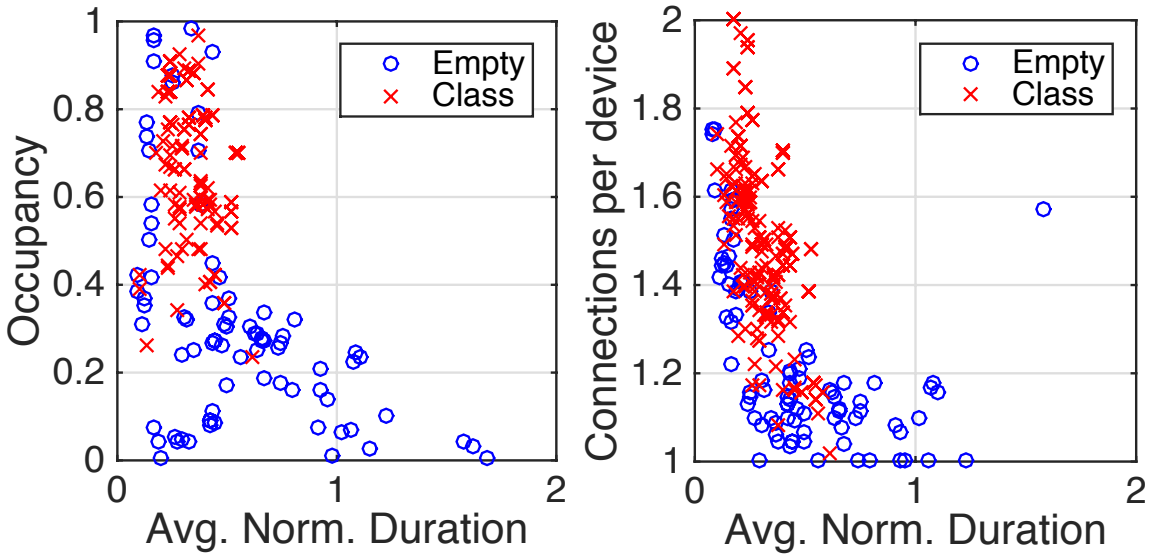


Fig. 5. The set of 213 non-empty time slots and 101 empty time slots plotted versus occupancy, average normalised duration and number of connections per device.

respectively. Such features capture the distribution of the connections starting instants. If such a distribution is sparse, it means that many devices started a connection in the same 5 minutes (something that typically happens during lecture breaks). If the distribution is uniform rather than sparse, it means that students tend to start connection randomly during the time slot and may indicate that they are not particularly "interested" or "focused" on the lecture.

Ideally, such features should be computed only for the connections started by devices who are actually located inside a classroom where a lecture is (or is not) given. Therefore, in the computation of the features we consider only those connections whose average signal quality during the selected time slot is greater than the average signal quality of all possible connections to the AP covering that classroom.

B. Rooms with classes VS. empty rooms

To analyse the differences between the behaviour of students during classes and empty slots, we use a classification-based approach. First, we partition the dataset into a training-set (80% of the time slots) and a test-set (the remaining 20%). The training set is used to train different classifiers based on (i) logistic regression, (ii) Linear Discriminant Analysis (LDA) and (iii)

TABLE I

CONFUSION MATRIX FOR THE EMPTY VS CLASS ROOM CLASSIFICATION

	Predicted: Empty	Predicted: Class
Actual: Empty	83.17%	16.83%
Actual: Class	1.88%	98.12%

Quadratic Discriminant Analysis (QDA), whose performance are evaluated on the test set. The same process is repeated 10 times, according to a stratified k-fold cross validation approach (i.e., maintaining the proportion of the different classes in each fold). Among the three classifiers, QDA exhibits the best performance and only its results are shown in the following for the sake of space.

Table I shows the performance of QDA in terms of a confusion matrix. As one can see, time slots in which a class was given are correctly classified 98% of the time, while this accuracy drops to 83% for empty time slots. Overall, the error rate (that is, the percentage of misclassified time slots) of the QDA classifier is as low as 8% and its F1-score is as high as 0.87. To further analyse what are the most discriminative features for such a classification, at each iteration of the k-fold validation we perform forward stepwise feature selection, each time keeping track of the top three selected features. It turns out that the three most discriminative features in this case are the *average normalised duration*, the *occupancy* and the *number of connections per device*. Figure 5 gives a visual explanation of the difference between empty and non-empty time slots: empty time slots are characterised by a lower average occupancy (0.6 vs. 0.9), and exhibit fewer connections per device (1.2 vs. 1.5) but with a longer normalized duration (0.6 vs. 0.3). That means that students occupying empty rooms tend to connect once to the wireless network and to maintain such a connection for a long time, while students attending a class connect more frequently but for shorter periods.

C. Engineering VS. architecture classes

We repeated the same approach for identifying differences between the behaviour of students during engineering or architecture courses. Table II shows the performance of QDA in terms of a confusion matrix. As one can see, even in this case the performance of classification is

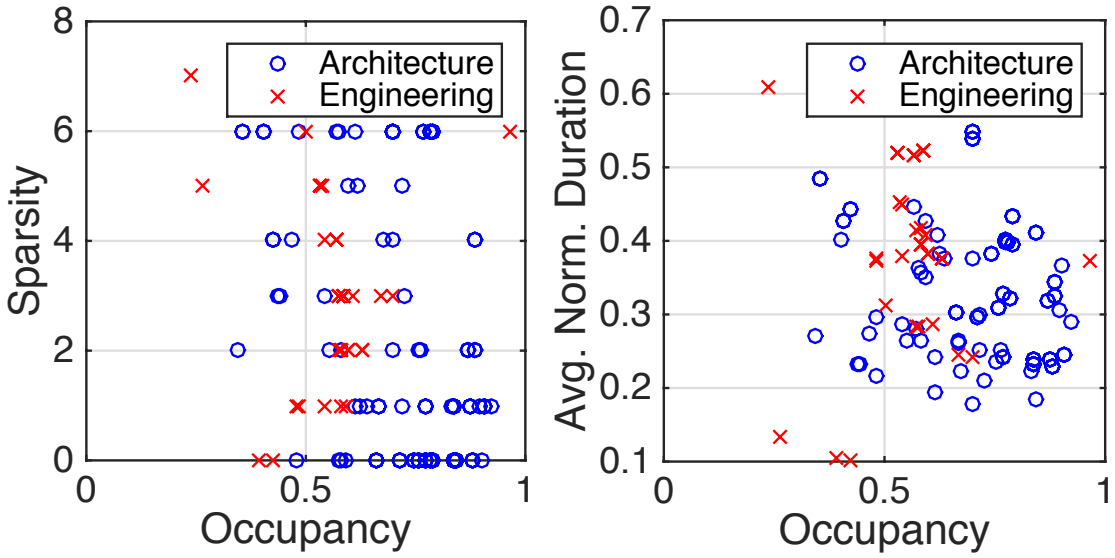


Fig. 6. The set of 156 architecture lectures and 57 engineering lectures plotted versus occupancy, normalised duration and sparsity of the connections distribution.

very good, with architecture and engineering classes being classified correctly 97% and 85% of the times, respectively. The overall error rate in this case is as low as 6% (F1-score equal to 0.96), meaning that there is indeed a difference in the usage of the network by architecture or engineering students and that the classifier is able to exploit such a difference. To better understand what changes between the behaviour of the two types of students, we again select the top three features obtained with stepwise forward selection, which turn out to be the *occupancy*, the *sparsity of the connection distribution* and the *average normalised duration*. Figure 6 shows the two types of classes when plotted on the planes individuated by such features. We can observe that engineering classes have on average (i) a lower occupancy (0.6 vs 0.8), (ii) a higher sparsity of the connection distribution (2.9 vs 1.9) and (iii) a higher normalised duration with respect to architecture students (0.35 vs 0.2). We do not have access to the exact number of enrolled students per class: assuming that such a number is proportional to the room capacity, our results seem to indicate that engineering students attend classes less frequently, but with more attention than architecture students.

TABLE II

CONFUSION MATRIX FOR THE ARCHITECTURE VS ENGINEERING LECTURES CLASSIFICATION

	Predicted: Architecture	Predicted: Engineering
Actual: Architecture	97.44%	2.56%
Actual: Engineering	14.04%	85.96%

V. CONCLUSIONS

The analysis of WiFi traffic traces can give fundamental insights on how to optimize and manage the network and, if coupled with other contextual information, it can also reveal patterns on how the end users behave. This high-level semantic information may be used as a basis for implementing future services, especially in the context of smart cities. In this work, we have analyzed traffic logs from a campus WiFi network. First, we have presented a spatio-temporal correlation analysis of the network under consideration. Then, we have used the traces to search for distinctive fingerprints in the logs themselves coming from different "types" of users; namely, we have proposed a supervised learning approach based on Quadratic Discriminant Analysis to classify WiFi traffic coming from empty or occupied classrooms and from engineering or architecture lectures. The proposed approach has been tested on a three-weeks WiFi data log with promising classification performance. Future works will address the exploitation of such results to implement high level services and applications.

REFERENCES

[1] P. Serrano, P. Salvador, V. Mancuso, and Y. Grunenberger, "Experimenting with commodity 802.11 hardware: Overview and future directions," *Communications Surveys Tutorials, IEEE*, vol. 17, no. 2, pp. 671–699, Secondquarter 2015.

[2] D. Kotz and K. Essien, "Analysis of a campus-wide wireless network," in *In Proceedings of ACM Mobicom*. ACM Press, 2002, pp. 107–118.

[3] T. Henderson, D. Kotz, and I. Abyzov, "The changing usage of a mature campus-wide wireless network," in *Proceedings of the Tenth Annual International Conference on Mobile Computing and Networking (MobiCom)*. ACM Press, September 2004, pp. 187–201. [Online]. Available: http://www.cs.dartmouth.edu/ dfk/papers/henderson-voice.pdf

[4] E. Zola and F. Barcelo-Arroyo, "A comparative analysis of the user behavior in academic wifi networks," in *Proceedings of the 6th ACM Workshop on Performance Monitoring and Measurement of Heterogeneous Wireless and Wired Networks*, ser. PM2HW2N '11. New York, NY, USA: ACM, 2011, pp. 59–66. [Online]. Available: http://doi.acm.org/10.1145/2069087.2069096

[5] F. Calabrese, J. Reades, and C. Ratti, "Eigenplaces: Segmenting space through digital signatures," *Pervasive Computing, IEEE*, vol. 9, no. 1, pp. 78–84, Jan 2010.

[6] M. Balazinska and P. Castro, "Characterizing mobility and network usage in a corporate wireless local-area network," in *Proceedings of the 1st International Conference on Mobile Systems, Applications and Services*, ser. MobiSys '03. New York, NY, USA: ACM, 2003, pp. 303–316. [Online]. Available: http://doi.acm.org/10.1145/1066116.1066127

[7] D. P. Blinn, T. Henderson, and D. Kotz, "Analysis of a Wi-Fi hotspot network," in *Proceedings of the International Workshop on Wireless Traffic Measurements and Modeling (WiTMeMo '05)*. USENIX Association, June 2005, pp. 1–6. [Online]. Available: http://www.cs.dartmouth.edu/ dfk/papers/blinn-hotspot.pdf

[8] M. Afanasyev, T. Chen, G. Voelker, and A. Snoeren, "Usage patterns in an urban wifi network," *Networking, IEEE/ACM Transactions on*, vol. 18, no. 5, pp. 1359–1372, Oct 2010.

[9] M. Afanasyev, T. Chen, G. M. Voelker, and A. C. Snoeren, "Analysis of a mixed-use urban wifi network: When metropolitan becomes neapolitan," in *Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC '08. New York, NY, USA: ACM, 2008, pp. 85–98. [Online]. Available: http://doi.acm.org/10.1145/1452520.1452531

[10] F. Ganji, Ł. Budzisz, F. G. Debele, N. Li, M. Meo, M. Ricca, Y. Zhang, and A. Wolisz, "Greening campus wlans: Energy-relevant usage and mobility patterns," *Computer Networks*, vol. 78, pp. 164 – 181, 2015, special Issue: Green Communications. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128614004241

[11] J. Yoon, B. D. Noble, M. Liu, and M. Kim, "Building realistic mobility models from coarse-grained traces," in *Proceedings of the 4th International Conference on Mobile Systems, Applications and Services*, ser. MobiSys '06. New York, NY, USA: ACM, 2006, pp. 177–190. [Online]. Available: http://doi.acm.org/10.1145/1134680.1134699

[12] M. Kim and D. Kotz, "Modeling users' mobility among wifi access points," in *Papers Presented at the 2005 Workshop on Wireless Traffic Measurements and Modeling*, ser. WiTMeMo '05. Berkeley, CA, USA: USENIX Association, 2005, pp. 19–24. [Online]. Available: http://dl.acm.org/citation.cfm?id=1072430.1072434

[13] M. Kim, D. Kotz, and S. Kim, "Extracting a mobility model from real user traces," in *INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings*, April 2006, pp. 1–13.

[14] J. Kim and A. Helmy, "Analysing the mobility, predictability and evolution of wlan users," *International Journal of Autonomous and Adaptive Communications Systems*, vol. 7, no. 1-2, pp. 169–191, 2014, pMID: 58020. [Online]. Available: http://www.inderscienceonline.com/doi/abs/10.1504/IJAACS.2014.058020

[15] A. Ruiz-Ruiz, H. Blunck, T. Prentow, A. Stisen, and M. Kjaergaard, "Analysis methods for extracting knowledge from large-scale wifi monitoring to inform building facility planning," in *Pervasive Computing and Communications (PerCom), 2014 IEEE International Conference on*, March 2014, pp. 130–138.

[16] T. S. Prentow, A. J. Ruiz-Ruiz, H. Blunck, A. Stisen, and M. B. Kjærgaard, "Spatio-temporal facility utilization analysis from exhaustive wifi monitoring," *Pervasive and Mobile Computing*, vol. 16, Part B, pp. 305 – 316, 2015, selected Papers from the Twelfth Annual {IEEE} International Conference on Pervasive Computing and Communications (PerCom 2014). [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1574119214001953

[17] U. Kumar, J. Kim, and A. Helmy, "Changing patterns of mobile network (wlan) usage: Smart-phones vs. laptops," in *Wireless Communications and Mobile Computing Conference (IWCMC), 2013 9th International*, July 2013, pp. 1584–1589.

[18] X. Wei, N. Valler, H. Madhyastha, I. Neamtiu, and M. Faloutsos, "A behavior-aware profiling of handheld devices," in *Computer Communications (INFOCOM), 2015 IEEE Conference on*, April 2015, pp. 846–854.

[19] U. Paul, A. Subramanian, M. Buddhikot, and S. Das, "Understanding traffic dynamics in cellular data networks," in

INFOCOM, 2011 Proceedings IEEE, April 2011, pp. 882–890.

Beam Based Stochastic Model of the Coverage Probability in 5G Millimeter Wave Systems

Cristian Tatino*†, Ilaria Malanchini†, Danish Aziz†, Di Yuan* *Department of Science and Technology, Linköping University, Sweden
Email: {cristian.tatino, di.yuan}@liu.se †Nokia Bell Labs, Stuttgart, Germany
Email: {ilaria.malanchini, danish.aziz}@nokia-bell-labs.com

Abstract

Communications using frequency bands in the millimeter-wave range can play a key role in future generations of mobile networks. By allowing large bandwidth allocations, high carrier frequencies will provide high data rates to support the ever-growing capacity demand. The prevailing challenge at high frequencies is the mitigation of large path loss and link blockage effects. Highly directional beams are expected to overcome this challenge. In this paper, we propose a stochastic model for characterizing beam coverage probability. The model takes into account both line-of-sight and first-order non-line-of-sight reflections. We model the scattering environment as a stochastic process and we derive an analytical expression of the coverage probability for any given beam. The results derived are validated numerically and compared with simulations to assess the accuracy of the model.

I. Introduction

The ever-growing data rate demand as well as the shortage of mobile frequency resources pose challenges for the upcoming fifth generation (5G) of mobile communications. A way to overcome these problems is to exploit unused frequency bands such as millimeter waves (mm-waves) between 30 to 300 GHz. Mm-waves bring new opportunities, but at the same time raise challenges, e.g., the large path loss caused by higher frequencies dramatically reduces the cell coverage area [1]. The use of highly directional narrow beams with high beamforming gain can help in increasing the cell coverage distance [2], but it requires robustness in procedures such as initial access, beam tracking, mobility management, and handovers.

The main focus of ongoing research related to mm-wave communications is the study of propagation characteristics, channel modeling, beam forming, and medium access control design.

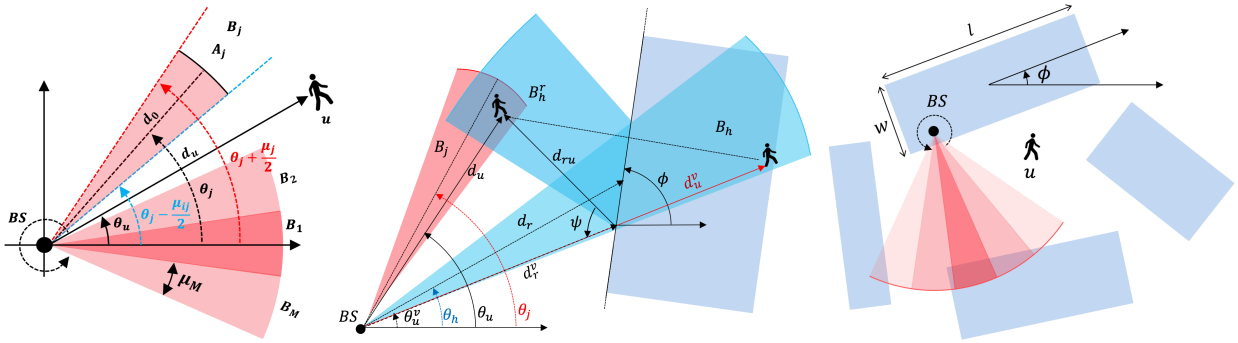
Extensive research is still needed to enable mm-wave communications to be deployed in cellular systems. To this end, we provide a beam based stochastic model for evaluating the coverage probability for any given beam. The analytical expression derived can be then exploited for supporting system level optimization, such as mobility management.

A. Related Works

Communications using mm-waves have been initially investigated for indoor and short range applications, where propagation is facilitated by line-of-sight (LOS) conditions and low-mobility. In [3], the authors propose two algorithms for beam searching, selection and tracking in wireless local area networks. They discretize the set of beams and find, by using iterative search, the best beam pair for the transmitter and the receiver. Similarly, the authors in [4] develop a method that compensates link blockage by switching between the LOS link and a non line-of-sight (NLOS) link, whenever the former is blocked. However, they do not provide any analytical model of the beam coverage and blockage probability.

Lately, the focus has shifted towards the application of mm-waves in outdoor scenarios and cellular systems. In [5], [6], the propagation characteristics of mm-waves are investigated. The study in [5] collects measurements taken in New York at 28 and 38 GHz. Results show that, when a high directional antenna array is used, path loss does not create a significant impediment to the propagation and it is still possible to reach the typical cell coverage of a high density urban environment. Based on the measurements reported in [5], [6] derives a statistical channel model for the path loss, the number of spatial clusters, the angular dispersion, and the outage probability.

Other works exploit stochastic geometry in order to derive statistical channel models and analyze the performance of mm-wave cellular systems. In [7], by proposing a stochastic model for the scattering environment, the authors compute the transmitter-receiver link blockage probability and the probability of coverage both for low frequency and mm-wave cellular networks. However, reflections are ignored. In [8], the authors propose an approach based on random shape theory to provide a statistical characterization of the mm-wave channel and to compute the power delay profile. The model takes into account both the LOS link and all the first-order reflections. Differently from our work, it considers omnidirectional antennas, at both the receiver and the transmitter, and it does not consider any beamforming approach. Leveraging the results in [7], a stochastic approach is adopted also in [9] to provide an analysis of the cell coverage probability



(a) (b) (c)

Fig. 1: (a) The BS and a subset of the beams that can be formed, indicated with different color shades. (b) Beam B_j covers directly the user u, whereas beam B_h covers the user u by the reflection B_h^r. (c) A mm-wave cell in an outdoor urban scenario with randomly deployed buildings.

and capacity. The authors demonstrate that, in high cell density conditions, mm-wave networks are able to provide sufficient signal-to-interference-plus-noise ratio (SINR) coverage and higher rate than the low-frequency cellular networks. Compared to previous works, they incorporate a directional beamforming for the SINR computation, at both the base station and the mobile user. However, in this case, reflections are ignored. The same assumption is done in [10], where the authors analyze the impact on the media access control layer design of highly directional communications for mm-waves. Using random shape theory, they investigate initial access and interference management, discussing handover and mobility issues.

B. Our Contributions

In this paper, we propose a beam based model that allows the analysis of the coverage probability and provides a useful tool to accurately investigate the effects of user mobility and beam selection in mm-wave cellular systems.

The main novelty, with respect to the state of the art, lies in the coupling of the following two aspects.

- The model incorporates beamforming, hence it allows to evaluate the coverage probability for any given beam, by taking into account the beam features (i.e., orientation and width) as well as the transmitter-receiver position.

- The model evaluates the coverage probability not only considering the direct beam but also including first-order reflections, which fairly contribute to the coverage probability in NLOS conditions [5].

The rest of the paper is structured as follows. Section II describes the system model and the assumptions. In Section III, we derive the beam coverage probability when blockages and first-order reflections are taken into account. In Section IV, we present a numerical evaluation to validate the accuracy of the model proposed. Section V concludes the paper.

II. System Model and Assumptions

We target the analysis of beam coverage probability in a cellular scenario. For any cell and user position of interest, the analysis deals with the coverage probability of any given beam. To this end, we consider a cell using mm-wave frequency bands for radio access. The base station (BS) is in the center of the cell and is equipped with a linear array of antennas that can form a discrete set of beams $\mathcal{M}$ of cardinality $|\mathcal{M}|$. Each beam B_j, $j \in \mathcal{M}$, is defined using a sector model and is fully specified by its direction θ_j and width μ_j, as shown in Fig. 1a. Furthermore, we assume the beamforming gain G_j to be a function of the beam width, i.e., $G_j = G(\mu_j)$. The generic user *u* is fully characterized by its position with respect to the BS, i.e., (θ_u, d_u), which is given in polar coordinates as shown in Fig. 1a, and is equipped with an omni-directional antenna.

In order to compute the *beam coverage probability*, we evaluate the signal-to-noise ratio (SNR) received by the generic user u from the BS, when a certain beam is used to transmit. Namely, in our model, we assume that the SNR depends on the distance d_u, the carrier frequency f, and the blockage effects caused by the scattering environment. Moreover, we evaluate the SNR considering either the LOS link or a first-order reflection, while excluding the contributions given by links with two or more reflections. This is motivated by the fact that beams reflected more than once arrive at the receiver with a very high path loss (caused by the longer path and a larger reflection loss) and therefore we assume their contributions to the SNR to be negligible. In particular, a first-order reflection B_j^r is generated when the beam B_j hits a building, as shown in Fig. 1b. We assume that beams are narrow enough to be totally reflected and we ignore diffraction and refraction effects. As a result, the beam B_j can cover the user either directly or by its first-order reflection B_j^r. In order to model the scattering environment, as shown in Fig. 1c, we consider buildings with rectangular shape. A building is specified by its center z, length l,

TABLE I: Summary of the notation

B_j	j^{th} beam
(μ_j, θ_j)	Width and orientation of B_j
A_j	Sector covered by B_j
B_j^r	Reflected beam generated by B_j
(θ_u, d_u)	Polar coordinates of u
(θ_u^v, d_u^v)	Polar coordinates of virtual user u
d_r	Distance between the BS and the obstacle along θ_h
d_r^v	Distance between the BS and the obstacle along θ_u^v
d_{ru}	Distance between the obstacle and u
$\text{SNR}_{B_j u}$	SNR for user u and beam B_j
$\text{SNR}_{B_j u}^D$	SNR for user u and direct beam B_j
$\text{SNR}_{B_j u}^R$	SNR for user u and reflected beam B_j^r
L, W, Φ	Length, width and orientation of an obstacle

width w and orientation ϕ. We assume all these to be independent random variables. Namely, the centers of the buildings Z form a homogeneous Poisson point process (PPP) of density λ. The lengths L and the widths W have probability density function $f_L(l)$ and $f_W(w)$, respectively. The orientations Φ are assumed to be uniformly distributed between $[0, \pi]$. A summary of the notation is reported in Table I.

III. Beam Coverage Probability

In this section, we compute the *beam coverage probability* of beam B_j, by explicitly considering the dependency between this probability and the beam properties (i.e., orientation and width) as well as the position of the user. We define the event $\mathsf{C}_{\mathsf{ju}} := \text{SNR}_{B_j u} \geq \Gamma$, where $\text{SNR}_{B_j u}$ is the SNR received by the user u for beam B_j and Γ is a given threshold. Formally, we define the *coverage probability* of beam B_j and the user u, i.e., $P(\mathsf{C}_{\mathsf{ju}})$, as:

$$P(\mathsf{C}_{\mathsf{ju}}) = P(\text{SNR}_{B_j u} \geq \Gamma). \tag{1}$$

In order to compute the SNR received by the user at the position (θ_u, d_u), we distinguish between two cases: the beam B_j covers the user directly or by a reflected beam B_j^r. Given the

assumption that a beam is either not reflected or totally reflected by a building, we consider those two events to be mutually exclusive, i.e., the probability that the same beam covers simultaneously the user both directly and with a reflection is set equal to zero. A beam B_j can cover the user directly if and only if the user u is inside the sector A_j, which is defined by the direction θ_j and the width μ_j as:

$$A_j = \left\{ (\theta, d) \in [0, 2\pi] \times [0, \infty] : |\theta - \theta_j| \leq \frac{\mu_j}{2} \right\}. \tag{2}$$

Therefore, we define the event $\mathsf{D_{ju}} := (\theta_u, d_u) \in A_j$ and we denote its complementary event as $\overline{\mathsf{D}}_{\mathsf{ju}}$. Thus, according to the law of total probability, $P(\mathsf{C_{ju}})$ can be written as:

$$P(\mathsf{C_{ju}}) = P(\mathsf{C_{ju}} \cap \mathsf{D_{ju}}) + P(\mathsf{C_{ju}} \cap \overline{\mathsf{D}}_{\mathsf{ju}}) = \\ P(\mathsf{C_{ju}}|\mathsf{D_{ju}})P(\mathsf{D_{ju}}) + P(\mathsf{C_{ju}}|\overline{\mathsf{D}}_{\mathsf{ju}})P(\overline{\mathsf{D}}_{\mathsf{ju}}), \tag{3}$$

where the first term of the sum represents the probability that the user is covered directly by the beam, while the second term is the probability to be covered by a reflection.

The first and the second addend of (3) are explicitly derived in Section III-A and in Section III-B, respectively.

A. Direct Beam Coverage Probability

The first term of (3) represents the probability of coverage with direct beam. According to the definition of $\mathsf{D_{ju}}$, we can write the probability $P(\mathsf{D_{ju}})$ as:

$$P(\mathsf{D_{ju}}) = \begin{cases} 1 & \forall\ (\theta_u, d_u) \in A_j \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Note that the event $\mathsf{D_{ju}}$ takes into account only whether the user lies in A_j (or not). In order to incorporate the blockage effect of obstacles, we define LOS_u (NLOS_u) as the event in which the user u is in LOS (NLOS) with respect to the BS. To compute the probability of LOS_u, we use one of the results derived in [7]. Namely, the authors show that (for the very same scattering model adopted here) the number of obstacles between the BS and the user is a random variable O that follows a Poisson distribution with mean:

$$E[O] = \beta d_u + p, \tag{5}$$

where $\beta = [2\lambda(E[L]+E[W])]/\pi,\ p = \lambda E[L]E[W]$ and $E[X]$ indicates the mean of the random variable X. Therefore, the probability that the user is in LOS can be written as follows:

$$P(\text{LOS}_u) = P(O = 0) = e^{-(\beta d_u + p)}. \tag{6}$$

Note that the two events $\mathsf{D_{ju}}$ and LOS_u are independent since the former depends only on θ_u and θ_j whereas the latter depends only on d_u. Furthermore, since LOS_u and NLOS_u are complementary events, the first term of (3) can be rewritten as follows:

$$P(\mathsf{C_{ju}} \cap \mathsf{D_{ju}}) = \\ P(\mathsf{C_{ju}}|\mathsf{D_{ju}} \cap \text{LOS}_u)P(\mathsf{D_{ju}})P(\text{LOS}_u) + \\ P(\mathsf{C_{ju}}|\mathsf{D_{ju}} \cap \text{NLOS}_u)P(\mathsf{D_{ju}})\left(1 - P(\text{LOS}_u)\right). \tag{7}$$

Moreover, by assumption, refraction is not considered in our model and a signal is completely reflected by an obstacle, hence $P(\mathsf{C_{ju}}|\mathsf{D_{ju}} \cap \text{NLOS}_u) = 0$.

Let $\{\text{SNR}_{B_j u}|(\mathsf{D_{ju}} \cap \text{LOS}_u)\}$ be the received SNR when user u is directly covered by beam B_j in LOS, which we indicate for the rest of the paper as $\text{SNR}^D_{B_j u}$. By applying the Friis' law we can write:

$$\text{SNR}^D_{B_j u} = \frac{P_t G_j G_u c^2}{(4\pi d_u f)^2 P_N}, \tag{8}$$

where P_t is the transmit power, c is the speed of light, G_u is the user beamforming gain, f is the frequency and P_N is the noise power.

To compute the coverage probability, we consider the case in which the SNR is greater than the given threshold Γ. Thus, let us define d_0 as the distance for which $\text{SNR}^D_{B_j u} = \Gamma$; with d_0 and A_j defining the *beam coverage area* as shown in Fig. 1a. Let $\mathbb{1}_{\mathcal{X}}(x)$ be the indicator function, i.e., $\mathbb{1}_{\mathcal{X}}(x) = 1\ \forall x \in \mathcal{X}$. We can then write the *direct beam coverage probability* as:

$$P(\mathsf{C_{ju}} \cap \mathsf{D_{ju}}) = P(\text{SNR}^D_{B_j u} \geq \Gamma)P(\mathsf{D_{ju}})P(\text{LOS}_u) = \\ \mathbb{1}_{[\theta_j - \frac{\mu_j}{2}, \theta_j + \frac{\mu_j}{2}] \times [0, d_0]}(\theta_u, d_u) e^{-(\beta d_u + p)}. \tag{9}$$

B. Reflected Beam Coverage Probability

We now investigate the probability of being covered by a first-order reflection. In general, B_j can generate different reflections, which depend on the position and orientation of the building

that is hit by the beam. In order to compute them, we assume the specular reflection law, i.e., the incident angle is assumed to be equal to the reflected one. Moreover, given the narrow beam assumption, we consider only the case in which the entire beam hits only one side of an obstacle (see Fig. 1b). Furthermore, the side of the building that is hit by the beam generates a straight line that divides the space in two half-planes, as shown in Fig. 1b. Thus, we compute the symmetric point (θ_u^v, d_u^v) of the user position with respect to this line, which we call *virtual user position*.

The user u is covered by a first-order reflected beam if the two events $\mathsf{R_{ju}}$ and LOS_u^R jointly hold, where $\mathsf{R_{ju}} := \{\theta_u^v \in A_j\} \cap \{d_u^v \geq d_r^v\}$ and LOS_u^R is the event in which the user u is in LOS with respect to the obstacle. Note that d_r^v is the distance between the BS and the obstacle along the direction identified by θ_u^v, as shown in Fig. 1b. By considering that the events $\overline{\mathsf{D}}_{\mathsf{ju}}$, $\mathsf{R_{ju}}$ and LOS_u^R are independent, we can write

$$P(\mathsf{C_{ju}}|\overline{\mathsf{D}}_{\mathsf{ju}}) = P(\text{SNR}_{B_j u}^R \geq \Gamma) P(\mathsf{R_{ju}}) P(\text{LOS}_u^R), \tag{10}$$

where $\text{SNR}_{B_j u}^R$ is the received SNR when beam B_j is reflected once by an obstacle. By applying the Friis' formula, we obtain

$$\text{SNR}_{B_j u}^R = \frac{P_t G_j^r G_u c^2}{(4\pi d_u^v f)^2 \sigma P_N}, \tag{11}$$

where the beamforming gain of the reflected beam is $G_j^r = G_j$ and σ is the reflection loss. Thus, we can derive the distance for which $\text{SNR}_{B_j u}^R = \Gamma$ as $d_0^v = \frac{d_0}{\sigma}$.

The events $\{\text{SNR}_{B_j u}^R \geq \Gamma\}$, $\mathsf{R_{ju}}$, and LOS_u^R and their respective probabilities depend on, e.g., d_u^v, θ_u^v, and d_{ru} (which is the distance between the user and the obstacle). Those in turn depends on the beam properties and the user position, which are both given, and on the distance of the first obstacle from the BS, d_r, and its orientation ϕ, see Fig. 1b. According to the stochastic model adopted for the scattering environment, those variables are described by probability density functions $f_{D_r}(d_r)$ and $f_\Phi(\phi)$, respectively. The latter is assumed to be uniformly distributed between $[0, \pi]$, whereas

$$f_{D_r}(d_r) = \delta(d_r)\left(1 - e^{-p}\right) + (1 - \delta(d_r))\,\beta e^{-(\beta d_r + p)} U(d_r), \tag{12}$$

where $\delta(r)$ is the Dirac delta function, i.e., $\delta(r) = 1$ for $r = 0$ and 0 otherwise, and $U(r)$ is the Heaviside step function. The details of the computation can be found in Appendix A.

$$P(\mathsf{C_{ju}} \cap \overline{\mathsf{D}}_{\mathsf{ju}}) = (1 - P(\mathsf{D_{ju}})) \times \int_o^{\pi} \int_0^{\infty} \mathbb{1}_{[\theta_j - \frac{\mu_j}{2}, \theta_j + \frac{\mu_j}{2}] \times [d_r^v, \frac{d_0}{\sigma}]}(\theta_u^v(r,\alpha), d_u^v(r,\alpha)) P(\mathbf{LOS}_u^R | r, \alpha) \left(\delta(r)\left(1 - e^{-p}\right) + (1 - \delta(r))\, \beta e^{-(\beta r + p)} U(r)\right) \frac{1}{\pi} \, \mathrm{d}r \, \mathrm{d}\alpha \tag{13}$$

To derive the final expression of the reflected beam coverage probability, reported in (13), we condition all terms of (10) on D_r and Φ. The product of the first two (conditioned) terms of (10) leads to the indicator function in (13), whereas the $P(\mathbf{LOS}_u^R | D_r = d_r, \Phi = \phi)$ is shown in Appendix B.

IV. Numerical Evaluation

In this section, we present the results of our study on the *beam coverage probability*. We assess the validity of our model by comparing the numerical results for $P(\mathsf{C_{ju}})$, computed using the analytical model, with simulation results. We used Matlab to compute numerically (13), hence (3), as well as to obtain the simulations results. It is important to note that in the simulation setup, we remove the assumption that a beam can hit only one side of an obstacle and we allow the beam to hit several obstacles (and sides), hence generating several reflections. Clearly, this makes the simulation environment more realistic, but also leads to some gap between the model and the simulation results, as shown later.

A. Simulation Setup

We consider a simulation area of 500×500 m^2 and we place the base station in the centre of the area. We independently generate $10,000$ instances by dropping the buildings randomly, according to a PPP of density λ. In order to obtain a comprehensive performance evaluation, hereafter, we vary several parameters, such as beam width and orientation, building density, and position of the user. The parameters that are fixed are: $P_t = 30$ dBm (as the experiments in [6]), $P_N = -85$ dBm, $f = 30$ GHz, and $\Gamma = 1$, i.e., 0 dB. W and L are characterized by uniform distribution between $[30, 50]$ and $[40, 60]$ (in meters), respectively. The reflection loss, which depends on several factors, e.g., angle of incidence on the obstacle, frequency, materials of the wall, is set to $\sigma = 3$ dB (as proposed in [8]), which means that half of the power is lost when the beam hits a building. Moreover, the results consider two different beam widths: $\mu_{ij} = 10^\circ$

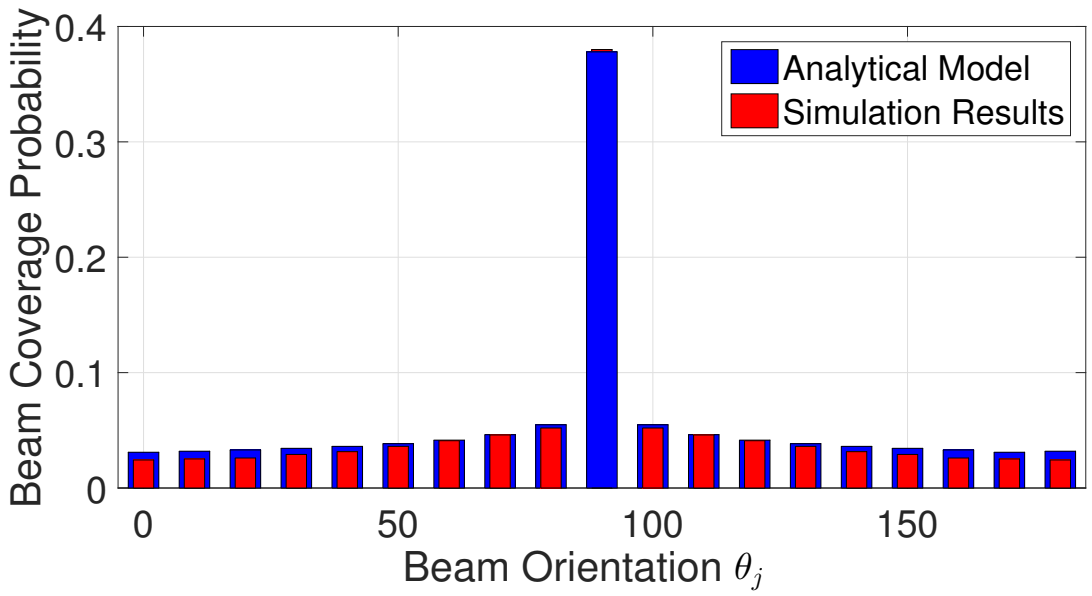


Fig. 2: The beam coverage probability computed for different non-overlapping beams with width $\mu_j = 10°$.

and $\mu_{ij} = 30°$. Since the gain depends on the beam width itself, we set $G(10°) = 36$ dBi and $G(30°) = 12$ dBi, which are assumed constant inside A_j and equal to 0 elsewhere. Moreover, we assume that $G_u = 1$ dBi.

B. Results

Fig. 2 shows the beam coverage probability when varying the beam orientation θ_j. The user position is $(90°, 50$ m$)$, the building density is $\lambda = 0.0002$ buildings/m^2, and the beam width is $\mu_j = 10°$. First, we observe that analytical and simulation results are very close to each other, validating the proposed model. Furthermore, we see that $P(\mathsf{C}_{\mathsf{ju}})$ decreases as the difference between the angular coordinate of the user, θ_u, and the beam direction, θ_j, increases. In particular, the coverage probability of the direct beam ($\theta_j = 90°$) is much larger than the ones obtained from reflected beams. Namely, when the beam direction moves away from the user angular coordinate, the path between the user and the obstacle becomes longer. Consequently, both the received SNR and the probability that the user is in LOS with respect to that obstacle decrease. Similar results have been obtained for different beam widths, but they are not reported for the sake of space.

Although the contribution to the coverage probability of the non-direct beams is smaller compared to the direct one, the aggregation of all of them can have a significant impact on the

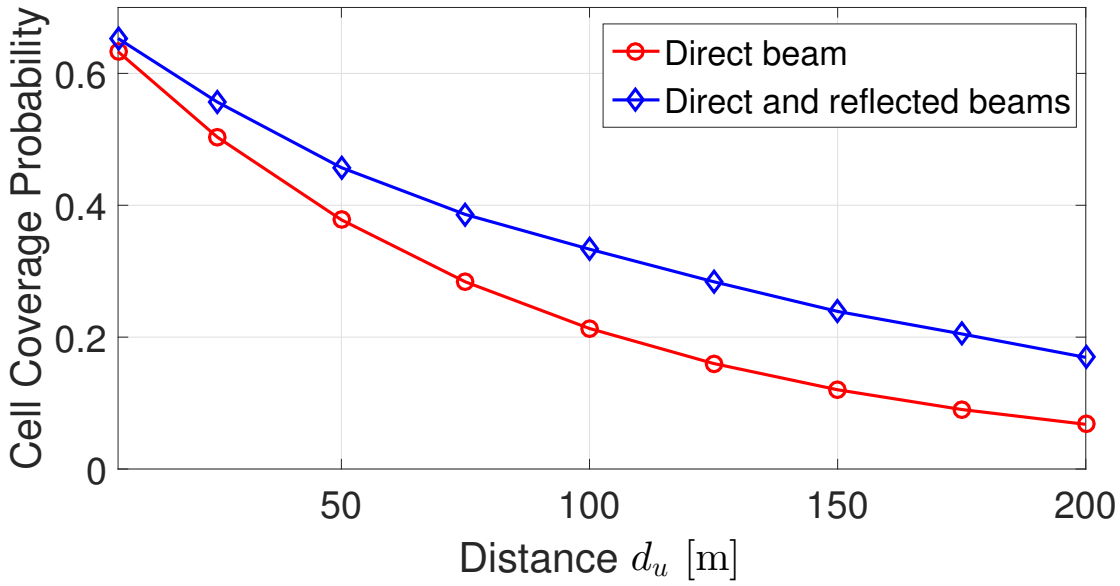


Fig. 3: Comparison between the cell coverage probability when considering only the direct beam and when including all possible reflected beams.

cell coverage probability. In Fig. 3, we compare the simulated cell coverage probability when only the direct beam is considered and when reflections are also included. Namely, in the latter case, the cell coverage probability is defined as the probability that at least one beam covers the user. We use the same parameters of Fig. 2, but we vary the user distance d_u. We observe that the cell coverage probability of the direct beam decreases faster in comparison to that with reflected beams. Moreover, the contribution of the reflections become more evident as the user distance increases. Similar conclusions can be drawn from Fig. 4, in which we show the beam coverage probability, varying the user distance d_u, for the direct beam and two reflected beams, when $\lambda = 0.0002$ buildings/m^2. In this case, the angular coordinate of the user is $\theta_u = 90°$ and the direct beam is the one with orientation $\theta_j = \theta_u$. Furthermore, we select two reflected beams: a first beam with $\theta_j = 95°$ and $\mu_j = 10°$, and a second beam with $\theta_j = 105°$ and $\mu_j = 30°$. In both cases, the reflected beam is chosen such that the user is placed at the border of the beam coverage angle, which corresponds to the best non-direct beam (in terms of beam coverage probability). We observe that the direct beam coverage probability decreases rapidly as d_u increases, whereas the reflected beam coverage probability remains almost constant. Fig. 4 further validates our model, reported with solid lines, with respect to simulations results, reported

with dashed lines. In particular, the divergence between model and simulation results, for the reflected beam curves, increases with d_u. This is due to the assumption (made in the analytical model, but not in the simulation setup) that the reflected beam hits only one side of the same building and is totally reflected, proving that our model is more accurate for narrow beams.

The validity of the model is shown also in Fig. 5, where we compare the beam coverage probability when the building density λ increases, for a fixed user position $(90°, 50 \text{ m})$ and the same beam set assumed in Fig. 4. In general, we observe that the analytical model matches well the simulation results. Moreover, the coverage probability of the direct beam decreases as the density λ increases, whereas the coverage probability of the reflected beams has a non-monotonic behavior. Namely, it increases from zero (when there are no buildings, hence no reflections) reaching a maximum for a given building density, and then decreases again. This behaviour is due to twofold effect that the building density has on the reflections. On one hand, increasing the building density corresponds to increasing the possibilities of generating reflections, which enhances the beam coverage probability. On the other hand, increasing the building density reduces the probability of LOS between the position of the first obstacle and the user, which decreases the beam coverage probability.

Finally, both Fig. 4 and Fig. 5 show that, by increasing the beam width, we can enhance the coverage probability of reflected beams (i.e., when there are NLOS conditions). This is due to the fact that the reflection of wider beams can cover a larger area and thus increase the probability of covering the user. This is an important outcome of our analysis, which suggests that width should be trade off between narrow beams, which are very good in LOS conditions, and wider beams, which can provide good coverage probability in NLOS conditions.

V. Conclusion

In this paper, we propose a beam based stochastic model for evaluating the beam coverage probability in mm-wave cellular systems. We model the scattering environment as a stochastic process and we derive an analytical expression valid for any given beam with respect to a given user position. The proposed model is able to capture the dependency of the beam coverage probability on various parameters, such as user position, beam orientation and width, and building density.

In general, the analytical model matches well the simulation results, especially for narrow beams. Furthermore, we show that, although the highest coverage probability is provided by

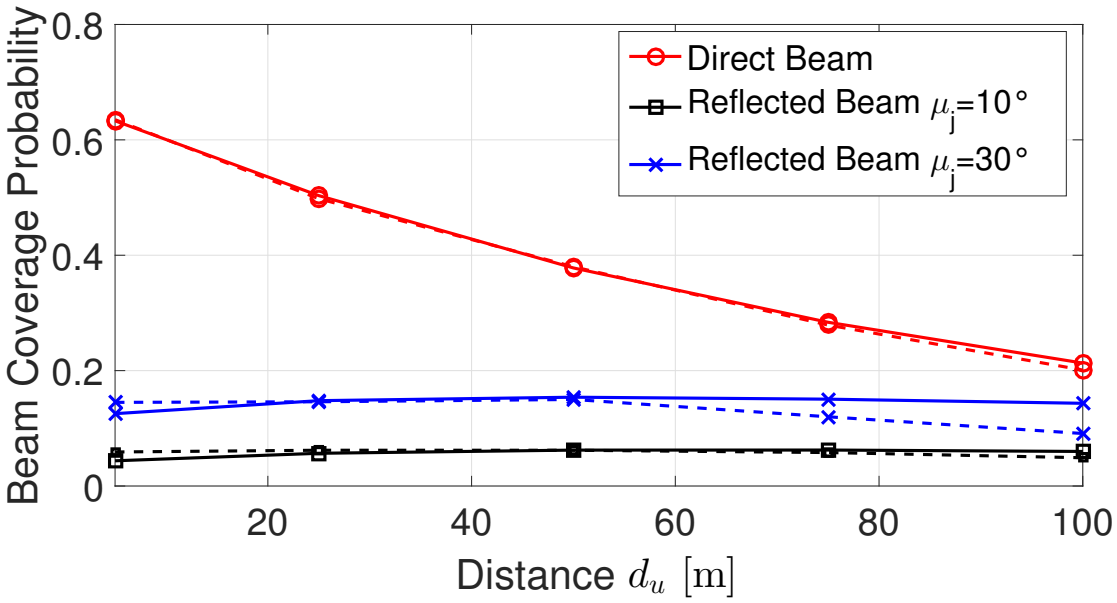


Fig. 4: Comparison between the analytical model (solid line) and simulation results (dashed line) of the direct and reflected beam coverage probability with varying user distance d_u.

the direct beam, reflections can fairly contribute to it, especially for larger user distances, i.e., when the LOS probability dramatically decreases. Moreover, the results show a non-monotonic behaviour of the reflected beam coverage probability with respect to the building density, which suggests that an optimal building density exists for NLOS conditions. Finally, we observe that increasing the beam width is a good strategy to improve the beam coverage probability in NLOS conditions.

Future work will further investigate the coverage properties due to reflections, and extend the model to improve the accuracy for wider beams. Furthermore, we will investigate how the proposed model can be used for network optimization, e.g., mobility management, in mm-wave systems.

VI. Appendix A

Hereafter, we derive the probability density function (PDF) of the distance of the first obstacle from the base station along a given direction, i.e., $f_{D_r}(d_r)$. Recall that the distribution of the total number of obstacles along a particular segment of distance d_r is a Poisson random variable

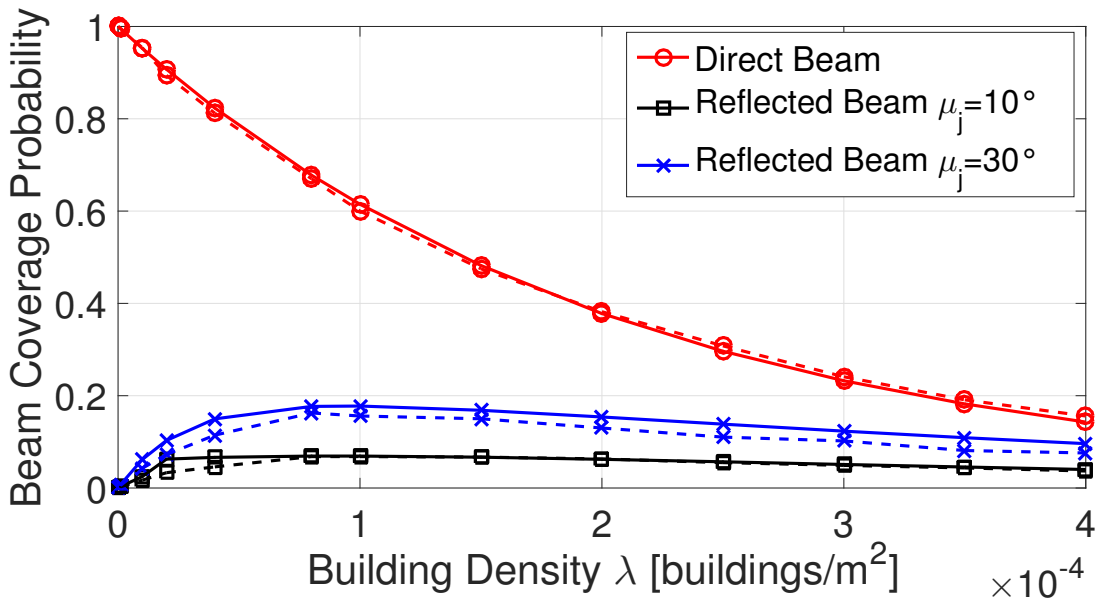


Fig. 5: Comparison between the analytical model (solid line) and simulation results (dashed line) of the direct and reflected beam coverage probability with varying building density λ.

O with mean defined in (5), cf. [7]. Therefore, the cumulative density function (CDF) $F_{D_r}(d_r)$ can be written as

$$F_{D_r}(d_r) = P(D_r \leq d_r) = 1 - P(D_r \geq d_r) = 1 - P(O(d_r) = 0) = 1 - e^{-(\beta d_r + p)}. \quad (14)$$

Since the CDF is equal to 0 for $d_r < 0$, it has a discontinuity in zero caused by the non-zero dimension of the obstacles. Therefore, to compute the PDF $f_{D_r}(d_r)$ we separate the two cases, i.e., $d_r = 0$ and $d_r > 0$. Then we obtain

$$f_{D_r}(d_r) = \begin{cases} 1 - P(O(0) = 0) = 1 - e^{-p} & d_r = 0 \\ \frac{\mathrm{d}F_{D_r}(d_r)}{\mathrm{d}d_r} = \beta e^{-(\beta d_r + p)} & d_r > 0. \end{cases} \quad (15)$$

VII. Appendix B

One can easily see that the event LOS_u^R is strongly correlated to the distance, d_r, between the BS and the first obstacle. For the sake of space we skip the details and we directly report the

derived approximation of $P(\text{LOS}_u^R | D_r = d_r, \Phi = \phi)$, which is

$$P(\text{LOS}_u^R | D_r = d_r, \Phi = \phi) = \begin{cases} \min(1, e^{-\beta d_{ru}+q}) & 0 \leq \psi \leq \frac{\pi}{2}, d_r \geq d_{ru} \\ \min(e^{-\beta(d_{ru}-d_r)}, e^{-\beta d_{ru}+q}) & 0 \leq \psi \leq \frac{\pi}{2}, d_r \leq d_{ru} \\ e^{-\beta d_{ru}} & \frac{\pi}{2} \leq \psi \leq \pi \end{cases} \quad (16)$$

where d_{ru} is the distance between the user and the obstacle, $q = \lambda \cot(\psi)(E[L^2] + E[W^2])/2$ and ψ is the angle formed by the reflection and directly depends on ϕ, as shown in Fig. 1b. $E[L^2]$, and $E[W^2]$ are the second moments of the length and width of the obstacles, respectively.

ACKNOWLEDGMENT

The authors would like to thank Dr. Vangelis Angelakis and Dr. Nikolaos Pappas for the insightful discussions.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 643002.

REFERENCES

[1] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: potentials and challenges," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366–385, March 2014.

[2] W. Roh, J. Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 106–113, February 2014.

[3] J. Wang, "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 8, pp. 1390–1399, October 2009.

[4] X. An, C.-S. Sum, R. Prasad, J. Wang, Z. Lan, J. Wang, R. Hekmat, H. Harada, and I. Niemegeers, "Beam switching support to resolve link-blockage problem in 60 GHz WPANs," in *IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications*, September 2009, pp. 390–394.

[5] T. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. Wong, J. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[6] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, June 2014.

[7] T. Bai, R. Vaze, and R. W. Heath, "Analysis of blockage effects on urban cellular networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 9, pp. 5070–5083, September 2014.

[8] N. A. Muhammad, P. Wang, Y. Li, and B. Vucetic, "Analytical model for outdoor millimeter wave channels using geometry-based stochastic approach," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 2, pp. 912–926, 2017.

[9] T. Bai and R. W. Heath, "Coverage and rate analysis for millimeter-wave cellular networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 2, pp. 1100–1114, February 2015.

[10] H. S. Ghadikolaei, C. Fischione, G. Fodor, P. Popovski, and M. Zorzi, "Millimeter wave cellular networks: A MAC layer perspective," *IEEE Transactions on Communications*, vol. 63, no. 10, pp. 3437–3458, October 2015.