A Survey of Anticipatory Mobile Networking: Context-Based Classification, Prediction Methodologies, and Optimization Techniques

Nicola Bui, *Student Member, IEEE*, Matteo Cesana, *Member, IEEE*, S. Amir Hosseini, *Student Member, IEEE*, Qi Liao, *Member, IEEE*, Ilaria Malanchini, *Member, IEEE*, and Joerg Widmer, *Senior Member, IEEE*

***Abstract*—A growing trend for information technology is to not just react to changes, but anticipate them as much as possible. This paradigm made modern solutions, such as recommendation systems, a ubiquitous presence in today's digital transactions. Anticipatory networking extends the idea to communication technologies by studying patterns and periodicity in human behavior and network dynamics to optimize network performance. This survey collects and analyzes recent papers leveraging context information to forecast the evolution of network conditions and, in turn, to improve network performance. In particular, we identify the main prediction and optimization tools adopted in this body of work and link them with objectives and constraints of the typical applications and scenarios. Finally, we consider open challenges and research directions to make anticipatory networking part of next generation networks.**

***Index Terms*—Anticipatory, prediction, optimization, 5G, mobile networks.**

I. Introduction

Evolving from one generation to the next, wireless networks have been constantly increasing their performance in many different ways and for diverse purposes. Among them, communication efficiency has always been paramount to increase the network capabilities without updating the entire infrastructure. This survey investigates anticipatory networking, a recent research direction that supports network optimization through system state prediction.

The core concept of anticipatory networking is that, nowadays, tools exist to make reliable prediction about network status and performance. Moreover, information availability is increasing every day as human behavior is becoming more socially and digitally interconnected. In addition, data centers are becoming more and more important in providing services and tools to access and analyze huge amounts of data.

As a consequence, not only can researchers tailor their solutions to specific places and users, but also they can anticipate the sequence of locations a user is going to visit or to forecast whether connectivity might be worsening, and to exploit the forecast information to take action before the event happens. This enables the possibility to take full advantage of good future conditions (such as getting closer to a base station or entering a less loaded cell) and to mitigate the impact of negative events (e.g., entering a tunnel).

This survey covers a body of recent works on anticipatory networking, which share two common aspects:

- *Anticipation:* they either explore prediction techniques directly or consider some future knowledge as given.
- *Networking:* they aim to optimize communications in mobile networks.

In addition, this survey delves into the following questions: How can prediction support wireless networks? Which type of information is possible to predict and which applications can take advantage of it? Which tools are the best for a given scenario or application? Which scenarios, among the ones envisioned for 5G networks, can benefit the most from anticipatory networking? What is yet to be studied in order for anticipatory networking to be implemented in 5G networks?

The main contributions of this survey are the following:

- A thorough **context-based analysis** of the literature classified according to the information exploited in the predictive framework.
- Two **handbooks on the prediction and optimization** techniques used in the literature, which allow the reader to get familiar with them and critically assess the different approaches.
- An analysis of the applicability of anticipatory networking techniques to different **types of wireless networks** and at different layers of the **protocol stack**.
- Summaries of all the main parts of the survey, highlighting **most popular choices and best practices**.

Manuscript received May 25, 2016; revised November 26, 2016 and March 21, 2017; accepted April 2, 2017. This work was supported in part by the European Union H2020-ICT (MONROE) under Grant 644399, in part by the European Union H2020-MSCA-ITN (ACT5G) under Grant 643002, in part by the Madrid Regional Government through the TIGRE5-CM Program under Grant S2013/ICE-2919, and in part by the Ramon y Cajal grant from the Spanish Ministry of Economy and Competitiveness under Grant RYC-2012-10788 and Grant TEC2014-55713-R. *(Corresponding author: Nicola Bui.)*

N. Bui and J. Widmer are with IMDEA Networks Institute, 28918 Madrid, Spain (e-mail: nicola.bui@imdea.org; joerg.widmer@imdea.org).

M. Cesana is with Politecnico di Milano, Milano, Italy (e-mail: matteo.cesana@polimi.it).

S. A. Hosseini is with the NYU Tandon School of Engineering, Brooklyn, NY 11201 USA (e-mail: amirhs.hosseini@nyu.edu).

Q. Liao and I. Malanchini are with Nokia Bell Labs, 70435 Stuttgart, Germany (e-mail: qi.liao@nokia-bell-labs.com; ilaria.malanchini@nokia-bell-labs.com).

Digital Object Identifier 10.1109/COMST.2017.2694140

AQ4

TABLE I
SURVEY CLASSIFICATION AND STRUCTURE

Context (Section III)		**Prediction** (Section IV)	**Optimization** (Section V)
	Geo	*Ideal:* [31, 42, 43, 45]	*ConvOpt[a]:* [43]
	Geo	*Time series:* [13, 28, 29, 32, 37, 38, 41]	*MDP[b]/MPC[c]:* [24, 26]
	Geo	*Regression, classification:* [14, 15, 22, 33-35, 44, 46]	*Game theory:* [131]
	Geo	*Probabilistic:* [11, 12, 16-21, 23-26]	*Heuristic:* [25, 32, 41, 42, 44-46]
	Link	*Ideal:* [56, 57, 65-70, 72-79]	*ConvOpt:* [64-70 72-79]
	Link	*Time series:* [54, 58, 59, 63]	*MDP/MPC:* [50, 60, 62, 158]
	Link	*Regression, classification:* [47-49, 51, 52, 55, 64]	*Game theory:* [129]
	Link	*Probabilistic:* [30, 50, 53, 60-62, 158]	*Heuristic:* [30, 47, 51, 54, 58, 59, 61, 63]
	Traffic	*Ideal:* [95-97, 111, 112, 115, 118, 138]	*ConvOpt:* [103-107, 111, 118-120, 138]
	Traffic	*Time series:* [100, 108-110, 113, 119, 145, 165]	*MDP/MPC:* [100, 115, 116, 165]
	Traffic	*Regression, classification:* [92-94, 98, 99, 101, 104-107, 114, 117, 156]	*Game theory:* [117]
	Traffic	*Probabilistic:* [93, 102, 116]	*Heuristic:* [96-99, 101, 112, 117]
	Social	*Ideal:* [121, 124, 137, 140]	*ConvOpt:* [126, 127, 137, 140, 159]
	Social	*Time series:* [40]	*MDP/MPC:* [157]
	Social	*Regression, classification:* [122, 123, 134, 139, 148, 149, 154]	*Game theory:* [128-131, 133]
	Social	*Probabilistic:* [125-127, 129, 130, 132, 135, 136, 157, 159]	*Heuristic:* [40, 121-125, 132, 148, 149]

[a] convex optimization [b] Markov decision process [c] model predictive control

- A final section analyzing **open challenges and potential issues** to the adoption of anticipatory networking solutions in future generation mobile networks.

A. Background and Guidelines

Anticipatory networking is the engineering branch that focuses on communication solutions that leverage the knowledge of the future evolution of a system to improve its operation. For instance, while a standard networking solution would answer the question *"which is the best user to be served?"*, an anticipatory equivalent would answer *"which are the best users to be served in the next time frames given the predicted evolution of their channel condition and service requirements?"*

A typical anticipatory networking solution is usually characterized by the following three attributes, which also determine the structure of this survey:

- *Context* defines the type of information considered to forecast the system evolution.
- *Prediction* specifies how the system evolution is forecast from the current and past context.
- *Optimization* describes how prediction is exploited to meet the application objectives.

To continue with the access selection example, the anticipatory networking solution might exploit the history of Global Positioning System (GPS) information (the *context*) to train an AutoRegressive (AR) model (the *prediction*) to predict the future positions of the users and their channel conditions to solve an Integer Linear Programming (ILP) problem (the *optimization*) that maximizes their Quality-of-Experience (QoE).

The main body of the anticipatory networking literature can be split into four categories based on the context used to characterize the system state and to determine its evolution: *geographic*, such as human mobility patterns derived from location-based information; *link*, such as channel gain, noise and interference levels obtained from reference signal feedback; *traffic*, such as network load, throughput, and occupied physical resource blocks based on higher-layer performance indicators; *social*, such as user's behavior, profile, and information derived from user-generated contents and social networks.

In order to determine which techniques are the most suitable to solve a given problem, it is important to analyze the following:

- *Properties* of the context:
  1) *Dimension* describes the number of variables predicted by the model, which can be uni- or multivariate.
  2) *Granularity and precision* define the smallest variation of the parameter considered by the context and the accuracy of the data: the lower the granularity, the higher the precision and vice versa. Temporal and spatial granularities are crucial to strike a balance between efficiency and accuracy.
  3) *Range* characterizes the distance (usually time or space) between known data samples and the farthest predicted sample. It is also known as prediction (or optimization) horizon.
- *Constraints* of the prediction or optimization model:
  1) *Availability of physical model* states whether a closed-form expression exists to describe the phenomenon.
  2) *Linearity* expresses the quality of the functions linking inputs and outputs of a problem.
  3) *Side information* determines whether the main context can be supported by auxiliary information.
  4) *Reliability and validity of information* specifies the noisiness of the data set, depending on which the prediction robustness should be calibrated.

The classification section will help the reader to understand the link between the different contexts and the solutions adopted to satisfy the given application requirements. Also, it is meant to provide a complete panorama of anticipatory networking. The two handbooks have the twofold objective of providing the reader with a short overview of the tools adopted in the literature and to analyze them in terms of variables of interest and constraints of the models. We believe that not only will this survey help researchers studying anticipatory networking, but also it will ease its adoption in future generation networks by providing a comprehensive overview

TABLE II
RELATED WORKS

Topic	**Content**
Big Data	[1] studies big data analytics for network optimization.
Context Information	[2], [3] discuss acquisition, modeling, exchange and usage of contextual information for different scenarios.
Data Classification	[4] surveys a variety of classifiers and uses them to predict unknown data.
Traffic & Throughput	[5] uses trace-driven simulation to compare prediction errors obtained using different techniques. [6] uses real network traffic to evaluate prediction techniques and to discuss their practical challenges.
Social Patterns	[7] uses social network information to study traffic patterns. [8] investigates the impact of prediction on QoE
Cognitive Radios	[9] investigates spectrum occupancy models and their reliability. [10] focus on spectrum occupancy and channel status prediction.

of research directions, available solutions and application scenarios.

Table I provides a mapping between the techniques described in Sections IV and V (columns) and the context discussed in Section III (rows). Each main category is further split into subcategories according to its internal structure. Namely, the prediction category is subdivided into ideal (perfect prediction is assumed to be available), time series predictive modeling, similarity-based classification and regression analysis, and probabilistic methods. The optimization category is split into Convex Optimization (ConvOpt), Markov Decision Process (MDP) and Model Predictive Control (MPC), game theoretic and, heuristic approaches.

The rest of the survey consists of a quick overview of other surveys on related topics in Section II, a context-based classification of the anticipatory networking literature in Section III, two handbooks on prediction and optimization techniques in Section IV and Section V, respectively. Sections VI and VII discuss how the anticipatory networking paradigm can be applied in a variety of network types and at different layers of the protocol stack. Sections VIII and VIII-C3 conclude the survey reporting the impact of anticipatory networking on future networks, the envisioned hindrances to its implementation and the open challenges.

II. RELATED WORK

This section discusses a few recent survey on topics close to anticipatory networking and is summarized in Table II.

Applying big data analytics for network optimization is studied in [1]. Based on the papers they reviewed, the authors propose a generic framework to support big data based optimization of mobile networks. Using traffic patterns derived from case studies, they argue that their framework can be used to optimize resource allocation, base station deployment, and interference coordination in such networks. In [2] and [3], the ability to extract and process contextual information by entities in a network is identified as a key factor in improving network performance. In [2], the procedure of using context information in wireless networks is broken down into acquisition, modeling, exchanging and evaluating stages, where the first two deal with gathering information and predicting the future behavior, and the latter two perform self-optimization and decision making. A similar taxonomy is provided in [3] and various examples of different techniques are reviewed for each phase. In addition to that, the authors provide a thorough survey on potential use cases of anticipatory networks and their respective challenges.

Predicting future states of network attributes is an essential task in designing anticipatory networks. Data classification, a popular prediction technique, has been thoroughly surveyed in [4]. Among other attributes, the prediction of data traffic and throughput has been the subject of [5] and [6]. Liu and Lee [5] consider seven algorithms for throughput prediction, ranging from mean-based and linear regression methods to Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) and compare their performance using a trace-driven simulator. Furthermore, they develop an information theoretic lower bound for the prediction error. In a similar attempt, [6] reviews real time Internet traffic classification. Here, the authors not only review prediction algorithms, but also try to shed light on practical challenges in deploying different kinds of techniques under different network scenarios. For instance, they argue that algorithms that require packet inspection either in the form of port number or payload, might have limited applicability due to potential encryption compared to methods that rely on statistical traffic properties.

The capability to extract user behavior in online social networks and use it to learn the evolution of traffic patterns in mobile networks is the subject of another survey [7]. The general approach of the papers included in that review is to use social graphs and classify different types of interactions between users on social networks in order to monitor the corresponding network traffic. Another important attribute for network performance is modeling the Quality of Experience (QoE) or how the service is perceived by the user. Baraković and Skorin-Kapov [8] provide a thorough survey including various methods for modeling QoE for different applications and also discuss tools for estimating and predicting QoE values by probing network parameters.

Cognitive Radio (CR) and Radio Environment Map (REM) are two very important technologies to measure, estimate and predict spectrum availability and occupancy. For instance, [9] and [10] provide two independent taxonomies of methodologies, campaigns and models. In addition, they review the reliability of these types of measurements [9] and they illustrate how to predict the system evolution thanks to available information and regression analysis [10].

To the best of our knowledge, this survey is the first to specifically address anticipatory techniques for mobile

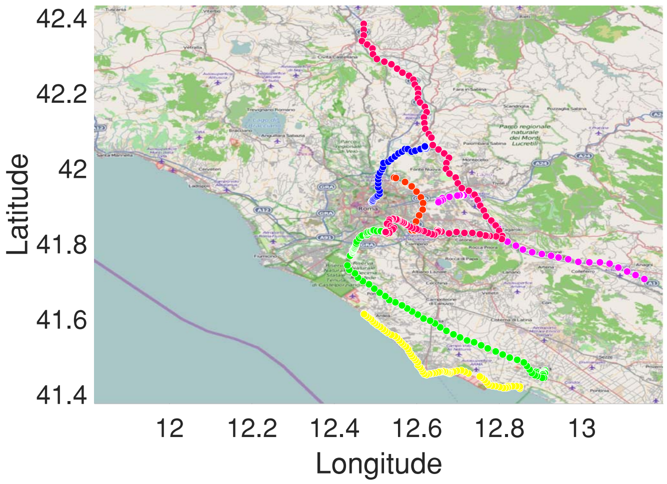


Fig. 1. Geographic context example: an example of estimated trajectories of 6 mobile users.

networks. We believe that, while the topic is undeniably hot, an overarching review of the body of work is still missing and greatly needed to facilitate the adoption of such a promising direction.

III. A Context-Based Classification of Anticipatory Networking Solutions

In this section, we classify the different types of context that can be predicted and exploited. For each one, we highlight the most popular prediction techniques as well as the applications for which an anticipatory optimization is performed.

A. Geographic Context

Geographic context refers to the geographic area associated to a specific event or information. In wireless communications, it refers to the location of the mobile users, often enriched with speed information as well as past and future trajectories. Understanding human mobility is an emergent research field that especially in the last few years has significantly benefited from the rapid proliferation of wireless devices that frequently report status and location updates. Fig. 1 illustrates an example of estimated trajectories of 6 mobile users.

The potential predictability in user mobility can be as high as 93% [11].[1] Along the same line, [12] investigates both the maximal predictability and how close to this value practical algorithms can come when applied to a large mobile phone dataset. Those results indicate that human mobility is very far from being random. Therefore, collecting, predicting and exploiting geographic context is of crucial importance.

In the rest of this section we organize the papers dealing with geographic context according to their main focus: the majority of them deals with pure geographical prediction and differs on secondary aspects such as whether they predict a single future location, a sequence of places or a trajectory. The second largest group of papers deals with multimedia streaming optimization.

1) Next Location Prediction: The simplest approach is to forecast where a given user will be at a predetermined instant of time in the future. Jiang *et al.* [13] propose to track mobile nodes using topological coordinates and topology preserving maps. Nodes' location is identified with a vector of distances (in hops) from a set of nodes called anchors and a linear predictor is used to estimate the mobile nodes' future positions. Evaluation is performed on synthetic data and nodes are assumed to move at constant speed. Results show that the proposed method approaches an accuracy above 90% for a prediction horizon of some tens of seconds.

A more general approach that exploits ANNs is discussed in [14]. Extreme Learning Machines (ELMs), which do not require any parameter tuning, are used to speed up the learning process. The method is evaluated using synthetic data over different mobility models.

To extend the prediction horizon [15] exploits users' locations and short-term trajectories to predict the next handover. The authors use Channel State Information (CSI) and handover history to solve a classification problem via supervised learning, i.e., employing a multi-class SVM. In particular, each classifier corresponds to a possible previous cell and predicts the next cell. A real-time prediction scheme is proposed and the feedback is used to improve the accuracy over time. Simulation results have been derived using both synthetic and real datasets. The longer moves along a given path, the higher the accuracy of forecasting the rest.

Location information can be extracted from cellular network records. In this way the granularity of the prediction is coarser, but positioning can be obtained with little extra energy. In particular, [16] aims at predicting a given user location from those of similar users. *Collective behavioral patterns* and a Markovian predictor are used to compute the next six locations of a user with a one-hour granularity, i.e., a six-hour prediction horizon. Evaluation is done using a real dataset and shows that an accuracy of about 70% can be achieved in the first hour, decreasing to $40-50\%$ for the sixth hour of prediction.

2) Space and Time Prediction: Prediction of mobility in a combined space-time domain is often modeled using statistical methods. In [17], the idea is to predict not only the future location a user will reach, but also *when* and for *how long* the user will stay there. To incorporate the *sojourn* time during which a user remains in a certain location, mobility is modeled as a semi-Markov process. In particular, the transition probability matrix and the sojourn time distribution are derived from the previous association history. Evaluation is done on a real dataset and shows approximately 80% accuracy. A similar approach is presented in [18], where the prediction is extended from single to multi-transitions (estimating the likelihood of the future event after an arbitrary number of transitions). Both papers provide also some preliminary results on the benefits of the prediction on resource allocation and balancing.

Barth *et al.* [19] represent the network coverage and movements using graph theory. The user mobility is modeled using a Continuous Time Markov (CTM) process where the prediction of the next node to be visited depends not only on

[1] Value obtained for a high-income country with stable social conditions. The percentage can decrease for different countries, e.g., low-income country or natural disaster situation.

the current node but also on the previous one (i.e., second-order Markovian predictor). Considering both local as well as global users' profiles, [20] extends the previous Markovian predictor and improves accuracy by about 30%. As pointed out in [21], sojourn times and transition probabilities are inhomogeneous. Thus, an inhomogeneous CTM process is exploited to predict user mobility. Evaluation on a real dataset shows an accuracy of 67% for long time scale prediction.

The interdependence between time and space is investigated also in [22] by examining real data collected from smartphones during a two-month deployment. Furthermore, [23] shows the benefit of using a location-dependent Markov predictor with respect to a location-independent model based on nonlinear time series analysis. Additionally, it is shown that information on arrival times and periodicity of location visits is needed to provide accurate prediction. A system design, named SmartDC, is presented in [24]–[26]. SmartDC comprises a mobility learner, a mobility predictor and an adaptive duty cycling. The proposed location monitoring scheme optimizes the sensing interval for a given energy budget. The system has been implemented and tested in a real environment. Notably, this is also one of the few papers that takes into account the *cost* of prediction, which in this case is evaluated in terms of energy. Namely, the authors detect approximately 90% of location changes, while reducing energy consumption at the expense of higher detection delay.

3) Location Sequences and Trajectories: A natural extension of the spatio-temporal perspective is the prediction of the location patterns and trajectories of the users. User mobility profiles have been introduced in [27] to optimize call admission control, resource management and location updates. Statistical predictors are used to forecast the next cell to which a mobile phone is going to connect. The validation of the solution is done via simulation. In [28], an approach for location prediction based on nonlinear time series analysis is presented. The framework focuses on the *temporal* predictability of users' location, considering their arrival and dwell time in relevant places. The evaluation is done considering four different real datasets. The authors evaluate first the predictability of the considered data and then show that the proposed nonlinear predictor outperforms both linear and Markov-based predictors. Precision approaches $70-90\%$ for medium scale prediction (5 minutes) and decreases to $20-40\%$ for long scale (up to 8 hours).

In order to improve the accuracy of time series techniques, De Domenico *et al.* [29] exploit the movement of friends, people, and, in general, entities, with correlated mobility patterns. By means of multivariate nonlinear time series prediction techniques, they show that forecasting accuracy approaches 95% for medium time scale prediction (5 to 10 minutes) and is approximately 50% for 3 hour prediction. Confidence bands show a significant improvement when prediction exploits patterns with high correlation. Evaluation is done considering two different real datasets.

Trajectory analysis and prediction also benefit from exploiting specific constraints such as streets, roads, traffic lights and public transportation routes. Fazio *et al.* [30] adapt the local Markovian prediction model for a specific coverage area in terms of a set of roads, moving directions, and traffic densities. When applying Markov prediction schemes, the authors consider a road compression approach to avoid dealing with a large number of locations, reduce the size of the state space, and minimize the approximation error. A more attractive candidate for trajectory prediction is the public transportation system, because of known routes and stops, and the large amount of generated mobile data traffic. Abou-Zeid *et al.* [31] investigate the predictability of mobility and signal variations along public transportation routes, to examine the viability of predictive content delivery. The analysis on a real dataset of a bus route, covering both urban and sub-urban areas, shows that modeling prediction uncertainty is paramount due to the high variability observed, which depends on combined effects of geographical area, time, forecasting window and contextual factors such as signal lights and bus stops.

Moving from discrete to continuous trajectories, Kalman filtering is used to predict the future velocity and moving trends of vehicles and to improve the performance of broadcasting [32]. The main idea is that each node should send the message to be broadcast to the fastest candidate based on its neighbors' future mobility. Simulation results show modest gains, in terms of percentage of packet delivery and end-to-end delay, with respect to non-predictive methods.

An alternative to Kalman filters is the use of regression techniques [33], which analyze GPS observations of past trips. A systematic methodology, based on geometrical structures and data-mining techniques, is proposed to extract meaningful information for location patterns. This work characterizes the location patterns, i.e., the set of locations visited, for several millions of users using nationwide call data records. The analysis highlights statistical properties of the typical covered area and route, such as its size, average length and spatial correlation.

Along the same line, [34] shows how the regularity of driver's behavior can be exploited to predict the current end-to-end route. The prediction is done by exploiting clustering techniques and is evaluated on a real dataset. A similar approach, named *WhereNext*, is proposed in [35]. This method predicts the next location of a moving object using past movement patterns that are based on both spatial and temporal information. The prediction is done by building a decision tree, whose nodes are the regions frequently visited. It is then used to predict the future location of a moving object. Results are shown using a real dataset provided by the GeoPKDD project [36]. The authors show the trade-off between the fraction of predicted trajectories and the accuracy. Both [34] and [35] show similar performance with an accuracy of approximately 40% and medium time scale prediction (order of minutes).

4) Dealing With Errors: The impact of estimation and prediction errors is modeled in [37]. The authors propose a comprehensive overview of several mobility predictors and associated errors and investigate the main error sources and their impact on prediction. Based on this, they propose a stochastic model to predict user throughput that accounts for uncertainty. The method is evaluated using synthetic data while assuming that prediction's errors have a truncated Gaussian distribution. The joint analysis on the predictability of location and signal strength, which in this case is simply quantified by the standard deviation of the random variable, shown in [31] indicates that location-awareness is a key factor to enable accurate signal strength predictions. Location errors are also considered in [38] where both temporal and spatial correlation are exploited to predict the average channel gain. The proposed method combines an AR model with functional linear regression and relies on location information. Results are derived using real data taken from the MOMENTUM project [39] and show that the proposed method outperforms SVM and AR processes.

5) Mobility-Assisted Handover Optimization: Seamless mobility requires efficient resource reservation and context transfer procedures during handover, which should not be sensitive to randomness in user movement patterns. To guarantee the service continuity for mobile users, the conventional in-advance resource reservation schemes make a bandwidth reservation over all the cells that a mobile host will visit during its active connection. With mobility pattern prediction, it is possible to prepare resources in the most probable cells for the moving users. Using a Markov chain-based pattern prediction scheme, Fazio *et al.* [30] propose a statistical bandwidth management algorithm to handle proactive resource reservations to reduce bandwidth waste. Along similar lines, [19], [40] investigate mobility prediction schemes, considering not only location information but also user profiles, time-of-day, and duration characteristics, to improve the handover performance in terms of resource utilization, handover accuracy, call dropping and call blocking probabilities.

6) Geographically-Assisted Video Optimization: One of the main applications that has been used to show the benefits of geographic context is video streaming. A pioneer work showing the benefit of a long-term location-based scheduling for streaming is [41]. The authors propose a system for bandwidth prediction based on geographic location and past network conditions. Specifically, the streaming device can use a GPS-based bandwidth-lookup service in order to predict the expected bandwidth availability and to optimally schedule the video playout. The authors present simulation as well as experimental results, where the prediction is performed for the upcoming 100 meters. The predictive algorithm reduces the number of buffer underruns and provides stable video quality.

Application-layer video optimization based on prediction of user's mobility and expected capacity, is proposed also in [42]–[44]. Lu and De Veciana [42] minimize a utility function based on system utilization and rebuffering time. For the single user case they propose an online scheme based on partial knowledge, whereas the multiuser case is studied assuming complete future knowledge. In [43], different types of traffic are considered: full buffer, file download and buffered video. Prediction is assumed to be available and accurate over a limited time window. Three different utility functions are compared: maximization of the network throughput, maximization of the minimum user throughput, and minimization of the degradations of buffered video streams. Both works show results using synthetic data and assuming perfect prediction of the future wireless capacity variations over a time window

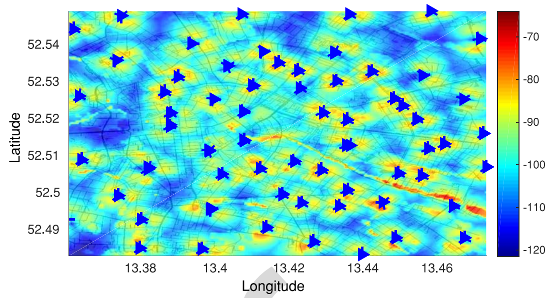


Fig. 2. Link context example: a pathloss map of Berlin downtown obtained from the data of the MOMENTUM project [39], where the triangles represent base stations. Pathloss maps are frequently used to predict the evolution of the connection quality in mobile networks.

with size ranging from tens to hundreds of seconds. In contrast, [44] introduces a data rate prediction mechanism that exploits mobility information and is used by an enhanced Proportionally Fair (PF) scheduler. The performance gain is evaluated using a real dataset and shows a throughput increase of 15%-55%.

Delay tolerant traffic can also benefit from offloading and prefetching as shown in [45]. The authors propose methods to minimize the data transfer over a mobile network by increasing the traffic offloaded to WiFi hotspots. Three different algorithms are proposed for both delay tolerant and delay sensitive traffic. They are evaluated using empirical measurements and assuming errors in the prediction. Results show that offloaded traffic is maximized when using prediction, even when this is affected by errors.

A *geo-predictive streaming system* called GTube, is presented in [46]. The application obtains the user's GPS locations and informs a server which provides the expected connection quality for future locations. The streaming parameters are adjusted accordingly. In particular, two quality adaptation algorithms are presented, where the video quality level is adapted for the upcoming 1 and n steps, respectively, based on the estimated bandwidth. The system is tested using a real dataset and shows that accuracy reaches almost 90% for very short time scale prediction (few seconds), but it decreases very fast approaching zero for medium time scale prediction (few minutes). However, the proposed n-step algorithm improves the stability of the video quality and increases bandwidth utilization.

B. Link Context

Link context refers to the prediction of the evolution of the physical wireless channel, i.e., the channel quality and its specific parameters, so that it is possible either to take advantage of future link improvements or to counter bad conditions before they impact the system. As an example of link context, Fig. 2 shows a pathloss map of the center of Berlin realized with the data of the MOMENTUM [39] project.

1) Channel Parameter Prediction: One possible approach to anticipate the evolution of the physical channel state is to predict the specific parameters that characterize it. In general, the variations of the physical channel can be caused by large-scale and small-scale fading. While predicting small-scale fading is quite challenging, if not impossible, several papers focuses on predicting path loss and shadowing effects. In [47], the time-varying nonlinear wireless channel model is adopted to predict the channel quality variation anticipating distance and pathloss exponent. The performance evaluation is done using both an indoor and an outdoor testbed. The good-put obtained with the proposed bitrate control scheme can be almost doubled compared to other approaches.

Pathloss prediction in urban environments is investigated in [48]. The authors propose a two-step approach that combines machine learning and dimensional reduction techniques. Specifically, they propose a new model for generating the input vector, the dimension of which is reduced by applying linear and nonlinear principal component analysis. The reduced vector is then given to a trained learning machine. The authors compare ANNs and SVMs using real measurements and conclude that slightly better results can be achieved using the ANN regressors.

Supporting the temporal prediction with spatial information is proposed in, e.g., [49] to study the evolution of shadow fading. The authors suggest to implement a Kriged Kalman Filter (KKF) to track the time varying shadowing using a network of CRs. The prediction is used to anticipate the position of the primary users and the expected interference and, consequently, to maximize the transmission rate of CR networks. Errors with the proposed model approach 2 dB (compared to 10 dB obtained with the pathloss based model). Targeting the same objective, but using a different methodology, [50] formulates the CR throughput optimization problem as an MDP. In particular, the predicted channel availability is used to maximize the throughput and to reduce the time overhead of channel sensing. Predictors robust to channel variations are investigated also in [51]. A clustering method with supervised SVM classification is proposed. The performance is shown for bulk data transport via Transport Control Protocol (TCP) and it is also shown that the predictive approach outperforms non-predictive ones.

Finally, maps can be used to summarize predicted information; for instance, algorithms to build pathloss maps are proposed in [52]. In this paper, the authors propose two kernel-based adaptive algorithms, namely the adaptive projected subgradient method and the multikernel approach with adaptive model selection. Numerical evaluation is done for both a urban scenario and a campus network scenario, using real measurements. The performance of the algorithms is evaluated assuming perfect knowledge of the users' trajectories.

2) Combined Channel and Mobility Context: Channel quality and mobility information are jointly predicted in [53]. The authors combine information on visited locations and corresponding achieved link quality to provide *connectivity forecast*. A Markov model is implemented in order to forecast future channel conditions. Location prediction accuracy is approximately 70% for a prediction window of 20 seconds. However, the location information has quite a coarse granularity (of about 100 m). In terms of bandwidth, the proposed model, evaluated on a real dataset, shows an accuracy within 10 KB/s for over 50% of the evaluation period, and within 50 KB/s for over 80% of the time. In [54], prediction is employed to adjust the routing metrics in ad hoc wireless networks. In particular, the metrics considered in the paper are the average number of retransmissions needed and the time expected to transmit a data packet. The solution anticipates the future signal strength using linear regression on the history of the link quality measurements. Simulations show that the packet delivery ratio is close to 100%, even though it drops to 20% using classical methods.

When the information used to drive the prediction is affected by errors, it is important to account for the magnitude of the error. This has been considered, for instance, in [55] and [56], where the impact of location uncertainties is taken into account. Namely, Muppirisetty *et al.* [55] show that classical Gaussian Process (GP) wrongly predicts the channel gain in presence of errors, while uncertain GP, which explicitly accounts for location uncertainty, outperforms the former in both learning and predicting the received power. Gains are shown also for a simple proactive resource allocation scenario. Similarly, Muppirisetty *et al.* [57] a proactive scheduling mechanism that exploits the statistical properties of user demand and channel conditions. Furthermore, the model captures the impact of prediction uncertainties and assesses the optimal gain obtained by the proactive resource scheduler. The authors also propose an asymptotically optimal policy that attains the optimal gain rapidly as the prediction window size increases. Uncertainties are also dealt with in [58], where a resource allocation algorithm for mobile networks that leverages link quality prediction is proposed. Time series filtering techniques (AutoRegressive and Moving Average (ARMA)) are used to predict near term link quality, whereas medium to long term prediction is based on statistical models. The authors propose a resource allocation optimization framework under imperfect prediction of future available capacity. Simulations are done using a real dataset and show that the proposed solution outperforms the limited horizon optimizer (i.e., when the prediction is done only for the upcoming few seconds) by $10-15\%$. Resource allocation is also addressed in [44], which extends the standard PF scheduler of 4G networks to account for data rate prediction obtained through adaptive radio maps.

3) Channel-Assisted Video Optimization: Wang *et al.* [59] propose an adaptive mobile video streaming framework, which stores video in the cloud and offers to each user a continuous video streaming adapted to the fluctuations of the link quality. The paper proposes a mechanism to predict the potential available bandwidth in the next time window (of a duration of a few seconds) based on the measurements of the link quality done in the previous time window. A prototype implementation of the proposed framework is used to evaluate the performance. This shows that the prediction has a relative error of about 10% for very short time windows (a couple of seconds) but becomes relatively poor for larger time windows. The video performance is evaluated in terms of "click-to-play" delay, which is halved with the proposed approach. A Markov model is used in [60], where information on both channel and buffer states is combined to optimize mobile video streaming. Both an optimal policy as well as a fast heuristic are proposed. A drive test was conducted to evaluate the performance of the proposed solution. In particular, the authors show the proportional dependency between utility and buffer size, as well as the complexity of the two algorithms. Furthermore, a Markov model is adopted to represent different user's achievable rates [61] and channel states [62]. The transition matrix is derived empirically to minimize the number of video stalls and their duration over a 10-second horizon.

Video calls are considered in [63]. Namely, a cross-layer design for proactive congestion control, named Rebera, is proposed. The system measures the real-time available bandwidth and uses a linear adaptive filter to estimate the future capacity. Furthermore, it ensures that the video sending rate never exceeds the predicted values, thereby preventing self-congestion and reducing delays. Performance results with respect to today's solutions are given for both a testbed and a real cellular network. Liu and Wei [64] propose a hop-by-hop video quality adaptation scheme at the router level to improve the performance of adaptive video streaming in Content Centric Networks (CCNs). In this context, the routers monitor network conditions by estimating the end-to-end bandwidth and proactively decrease the video quality when network congestion occurs. Performance is evaluated considering a realistic large-scale network topology and it is shown that the proposed solution outperforms state of the art schemes in terms of both playback quality and average delay.

4) Video Optimization Under Uncertainty: For the video optimization use case, some works also assess the impact of uncertain predictions. Dräxler *et al.* [65] propose a stochastic model of prediction errors, based on [37], and introduce an online scheduler that is aware of prediction errors. Namely, based on the expected prediction accuracy, the algorithm determines whether to consider or discard the predicted data rate. A similar model for prediction errors is introduced in [66]. In this case, a Linear Programming (LP) formulation is proposed to trade off spectral efficiency and stalling time. The proposed solution shows good gains with respect to the case without prediction, even when errors occur. LP is used also in [67] to minimize the base station airtime with the constraint of no video interruption. In this case, uncertainties are modeled by using a fuzzy approach. Furthermore, in order to keep track of the previous values of the error, a Kalman filter is used. Simulations are run using synthetic data and show the effect of channel variability on video degradation and average airtime. In [68], bandwidth prediction is exploited to increase the quality of video streaming. Both perfect and uncertain prediction are considered and a robust heuristic is proposed to mitigate the effect of prediction errors when adapting the video bitrate. In [69] and [70], a predictive resource allocation robust to rate uncertainties is proposed. The authors propose a framework that provides quality guarantees with the objective of minimizing energy consumption. Both optimal gradient-based and real-time guided heuristic solutions are presented. In [69] both Gaussian and Bernstein approximation are used to model rate uncertainties, whereas [70] considers only the former one. Similarly, [71] provides predictive Quality-of-Service (QoS) over wireless Asynchronous Transfer Mode (ATM) networks: given the TDMA nature of these networks, these schemes optimize the number of allocated time slots depending on the characteristics of the traffic stream and the wireless link.

5) Efficiency Bounds and Approximations for Multimedia Streaming Applications: A few papers [72]–[79] investigate resource allocation optimization assuming that the future channel state is perfectly known. While addressing different objectives, these papers share similar methods: they first devise a problem formulation from which an optimal solution can be obtained (using standard optimization techniques), then they propose sub-optimal approaches and on-line algorithms to obtain an approximation of the optimal solution. Furthermore, all these papers leverage a buffer to counteract the randomness of the channel. For instance, in case a given amount of information has to be gathered within a deadline, the buffer allows the system to optimize (for a given objective function) the resource allocation while meeting the deadline.

In this regard, energy-efficiency is the primary objective in [72] and [73], which is optimized by allowing the network base stations to be switched off once the users' streaming requirements have been satisfied. Simulations show that an energy saving up to 80% with respect to the baseline approach can be achieved and that the performance of the heuristic solution is quite close to the optimal (but impractical) Mixed-Integer Linear Programming (MILP) approach. Buffer size is investigated in [78], where the author introduces a linear formulation that minimizes the amount for resources assigned to non-real time video streaming with constraints on the user's playout buffer. Results are shown for a scenario with both video and best effort users and highlight the gain in terms of required resources to serve the video users as well as data rate for the best effort users.

The trade-off between streaming interruption time and average quality is investigated in [76] and [77] by devising a mixed-integer quadratically constrained problem which computes the optimal download time and quality for video segments. Then, the authors propose a set of heuristics tailored to greedily optimize segment scheduling according to a specific objective function, e.g., maximum quality, minimum streaming interruption, or fairness. Similar objectives are tackled in [74] and [75] in a lexicographic approach, so that streaming continuity is always prioritized over quality. They first propose a heuristic for the lateness-quality problem that performs almost as good as the MILP formulation. Then, they extend the MILP formulation to include QoS guarantees and they introduce an iterative approximation based on a simpler LP formulation. A further heuristic approach is devised in [79] and accounts for the buffer and channel state prediction. The proposed approach maximizes the streaming quality while guaranteeing that there are no interruptions.

6) Cognitive Radio Maps: CRs are context-aware wireless devices that adapt their functionalities to changes in the environment. They have been recently used [80]–[82] to obtained the so-called REM: a multi-dimensional database containing a wide set of information ranging from regulations to spectrum usage.

For instance, REM are used to predict spectrum availability in CR [80]: the paper exploits cognitive maps to provide contextual information for predictive machine learning approaches such as Hidden Markov Models (HMM), ANN and regression techniques. The construction of these maps is discussed in [81] and the references therein, while their use as enabler for CR networks is analyzed in [82].

In the context of anticipatory networking, REMs are often used as a source of contextual information for the actual prediction technique adopted, rather than as prediction tools themselves. References [9] and [10] present two surveys of methodologies and measurement campaigns of spectrum occupancy. In particular, [9] proposes a conservative approach to account for measurement uncertainty, while [10] exploits predictors to provide the future channel status. In addition, prediction through machine learning approaches is addressed in [83], where different techniques are compared to assess future channel availability.

Imperfect measurements are dealt with in [84], which models the problem as a repeated game and maximizes the total network payoff. However, in cognitive networks, the channel status depends on the activity of primary users. Reference [85] surveys the models proposed so far to describe primary users activity and that can be used to drive prediction in this area. Once the activity of primary users is available or predicted, it is possible to control the activity of secondary users in order to guarantee the agreed QoS to the former [86], [87]. These papers compute the feasible cognitive interference region in order to allow secondary users' communication respecting primary users' rights. The utilization of spectrum opportunity describes the probability of a secondary user to exploit a free communication slot [88].

A similar form of opportunistic spectrum usage goes under the name of white space [89]: i.e., channels that are unused at specific location and time. CRs can take advantage of these frequencies thanks to dynamic spectrum access. Finally, [90] describes how to exploit CR to realize a complete smart grid scenario; [91] describes how to exploit channel bonding to increase the bandwidth and decrease the delay of CR.

C. Traffic Context

This section overviews some of the approaches that focus on traffic and throughput prediction. Although related to the previous context, the papers discussed in this section leverage information collected from higher layers of the protocol stack. For instance, solutions falling in this category try to predict, among other parameters, the number of active users in the network and the amount of traffic they are going to produce. Similarly, but from the perspective of a single user, the prediction can target the data rate that a streaming application is going to achieve in the near term.

We grouped these papers in three main classes: pure analysis of mobile traffic; traffic prediction for networking optimization; and direct throughput prediction.

1) Traffic Analysis and Characterization: The analysis of mobile traffic is fundamental for long-term network optimization and re-configuration. To this end, several pieces of work have addressed such research topics in the recent past.

The work in [92] targets the creation of regressors for different performance indicators at different spatio-temporal granularity for mobile cellular networks. Namely, the authors focus on the characterization of per-device throughput, base station throughput and device mobility. A one-week nationwide cellular network dataset is collected through proprietary traffic inspection tools placed in the operator network and are used to characterize the per-user traffic, cell-aggregate traffic and to perform further spatio-temporal correlation analysis.

A similar scope is addressed by [93] which, on the other hand, focuses more on core network measurements. Flow level mobile device traffic data are collected from a cellular operator's core network and are used to characterize the IP traffic patterns of mobile cellular devices.

More recently, Sayeed *et al.* [94] studied traffic prediction in cloud analytics and prove that optimizing the choice of metrics and parameters can lead to accurate prediction even under high latency. This prediction is exploited at the application/TCP layer to improve the performance of the application avoiding buffer overflows and/or congestion.

2) Traffic Prediction: Several applications can benefit from the prediction of traffic performance features. For instance, a predictive framework that anticipates the arrival of upcoming requests is used in [95] to prefetch the needed content at the mobile terminal. The authors propose a theoretical framework to assess how the outage probability scales with the prediction horizon. The theoretical framework accounts for prediction errors and multicast delivery. Along the same line, queue modeling [96] and analysis [97] is used to predict the upcoming workloads in a lookahead time window. Leveraging the workload prediction, a multi-slot joint power control and scheduling problem is formulated to find the optimal assignment that minimizes the total cost [96] or maximizes the QoS [97].

Multimedia optimization is the focus in [98]. By predicting throughput, packet loss and transmission delay half a second in advance, the authors propose to dynamically adjust application-level parameters of the reference video streaming or video conferencing services including the compression ratio of the video codec, the forward error correction code rate and the size of the de-jittering buffer. Traffic prediction is also addressed in [99], where the authors propose to use a database of events (concerts, gatherings, etc.) to improve the quality of the traffic prediction in case of unexpected traffic patterns and in [100], where a general predictive control framework along with Kalman filter is proposed to counteract the impact of network delay and packet loss. The objective of [101] is to build a model for user engagement as a function of performance metrics in the context of video streaming services. The authors use a supervised learning approach based on average bitrate, join time, buffering ratio and buffering to estimate the user engagement. Finally, inter-download time can be modeled [102] and subsequently predicted for quality optimization.

The work in [103] targets energy-efficient resource scheduling in mobile radio networks. The authors introduce a Mixed Non-Linear Program (MNLP) which returns on a slot basis the optimal allocation of resources to users and the optimal users-cell association pattern. The proposed model leverages optimal traffic predictors to obtain the expected traffic conditions in the following slots. Radio resource allocation in mobile radio networks is addressed also in [104] and later by the same authors in [105]; the target is to design a predictive framework to optimally orchestrate the resource allocation and network selection in case one operator owns multiple access networks. The predictive framework aims at minimizing the expected time average power consumption while keeping the network (user queues) stable. The core contribution of [106] and [107] is the use of deep learning techniques to predict the upcoming video traffic sessions; the prediction outcome is then used to proactively allocate the resources of video servers to these future traffic demands.

3) Throughput Prediction: Rather than predicting the expected traffic or optimizing the network based on traffic prediction, the work in this section targets the prediction/optimization based on the expected throughput. A common characteristic of the work described here is that the spatio-temporal correlation is exploited in the prediction phase of the expected throughput.

Quite a few early works studied how to effectively predict the obtainable data rate. In particular, long term prediction [108] with 12-hour granularity allows to estimate aggregate demands up to 6 months in advance. Shorter and variable time scales are studied in [109] and [110] adopting AutoRegressive Integrated and Moving Average (ARIMA) and Generalized AutoRegressive Conditionally Heteroskedastic (GARCH) techniques.

Abou-Zeid and Hassanein [111] propose a dynamic framework to allocate downlink radio resources across multiple cells of 4G systems. The proposed framework leverages context information of three types: radio maps, user's location and mobility, as well as application-related information. The authors assume that a forecast of this information is available and can be used to optimize the resource allocation in the network. The performance of the proposed solution is evaluated through simulation for the specific use case of video streaming. Geo-localized radio maps are also exploited in [112]. Here the optimization is performed at the application layer by letting adaptive video streaming clients and servers dynamically change the streaming rate on the basis of the current bandwidth prediction from the bandwidth maps. The empirical collection of geo-localized data rate measures is also addressed in [113] which introduces a dataset of adaptive Hypertext Transfer Protocol (HTTP) sessions performed by mobile users.

The work in [114] considers the problem of predicting end-to-end quality of multi-hop paths in community WiFi networks. The end-to-end quality is measured by a linear combination of the expected transmission count across all the links composing the multi-hop path. The authors resort to a real data set of a WiFi community network and test several predictors for the end-to-end quality.

The anticipation of the upcoming throughput values is often applied to the optimization of adaptive video streaming services. In this context, Yin *et al.* [115] leverage throughput prediction to optimally adapt the bit rate of video encoders; here, prediction is based on the harmonic mean of the last k throughput samples.

Sun *et al.* [116] and Jiang *et al.* [117] build on the conjecture that video sessions sharing the same critical features have similar QoE (e.g., re-buffering, startup latency, etc.). Consequently, first clustering techniques are applied to group similar video sessions, and then throughput predictors based on HMMs are applied to each cluster to dynamically adapt the bit rate of the video encoder to the predicted throughput samples.

The work in [118] resorts to a model-based throughput predictor in which the throughput of a Dynamic Adaptive Streaming over HTTP (DASH)-based video streaming service is assumed to be a random variable with Beta-like distribution whose parameters are empirically estimated within an observation time window. Building on this estimate, the authors propose a MNLP with a concave objective function and linear constraints. The program is implemented as a multiple choice knapsack problem and solved using commercial solvers. Along the same lines, the optimization of a DASH-based video streaming service is addressed in [119], where the authors propose an adaptive video streaming framework based on a smoothed rate estimate for the video sessions.

The work in [120] considers the scenario where a small cell is used to deliver video content to a highly dense set of users. The video delivery can also be supported in a distributed way by end-user devices storing content locally. A control-theoretic framework is proposed to dynamically set the video quality of the downloaded content while enforcing stability of the system.

D. Social Context

The work on anticipatory networking leveraging social context exploits *ex ante* or *ex post* information on social-type relationships between agents in the networking environment. Such information may include: the network of social ties and connections, the user's preference on contents, measures on user's centrality in a social network, and measures on users' mobility habits. The aforementioned context information is leveraged in three main application scenarios: caching at the edge of mobile networks, mobility prediction, and downlink resource allocation in mobile networks.

1) Social-Assisted Caching: Motivated by the need of limiting the load in the backhaul of 5G networks, references [121]–[123] propose two schemes to proactively move contents closer to the end users. In [121], caching happens at the small cells, whereas in [122] and [123] contents can be proactively downloaded by a subset of end users which then re-distribute them via device-to-device (D2D) communication. The authors first define two optimization problems which target the load reduction in the backhaul (caching at small cells) and in the small cell (caching at end users), respectively, then heuristic algorithms based on machine learning tools are proposed to obtain sub-optimal solutions in reasonable processing time. The heuristic first collects users' content rating/preferences to predict the popularity matrix $\mathbf{P}_m$. Then, content is placed at each small cell in a greedy way starting from the most popular ones until a storage budget is hit. The first algorithmic step of caching at the end users is to identify the *K* most connected users and to cluster the remaining ones in communities. Then it is possible to characterize the content preference distributions within each community and greedily place contents at the cluster heads. In [123], the prediction leverages additional information on the underlying structure of content popularity within the communities of users. Joint mobility and popularity prediction for content caching at small cell base stations is studied in [124]. Here, the authors propose a heuristic caching scheme that determines whether a particular content item should be cached at a particular base station by jointly predicting the mobility pattern of users that request that item as well as its popularity, where popularity prediction is performed using the inter-arrival times of consecutive requests for that object. They conclude that the joint scheme outperforms caching with only mobility and only popularity models.

A similar problem is addressed in [125]: the authors consider a distributed network of femto base stations, which can be leveraged to cache videos. The authors study where to cache videos such that the average sum delay across all the end users is minimized for a given video content popularity distribution, a given storage capacity and an arbitrary model for the wireless link. A greedy heuristic is then proposed to reduce the computational complexity.

In [126] and [127], it is argued that proactive caching of delay intolerant content based on user preferences is subject to prediction uncertainties that affect the performance of any caching scheme. In [126], these uncertainties are modeled as probability distributions of content requests over a given time period. The authors provide lower bounds on the content delivery cost given that the probability distribution for the requests is available. They also derive caching policies that achieve this lower bound asymptotically. It is shown that under uniform uncertainty, the proposed policy breaks down to equally spreading the amount of predicted content data over the horizon of the prediction window. Another approach to solve the same problem is used in [127], where personalized content pricing schemes are deployed by the service provider based on user preferences in order to enhance the certainty about future demand. The authors model the pricing problem as an optimization problem. Due to the non-convex nature of their model, they use an iterative sub-optimal solution that separates price allocation and proactive download decisions.

2) Social-Assisted Matching Game Theory: Matching game theory [128] can be used to allocate networks resources between users and base stations, when social attributes are used to profile users. For instance, by letting users and base stations rank one another to capture users' similarities in terms of interests, activities and interactions, it is possible to create social utility functions controlling a distributed matching game. In [129], a self-organizing, context-aware framework for D2D resource allocation is proposed that exploits the likelihood of strongly connected users to request similar contents. The solution is shown to be computationally feasible and to offer substantial benefits when users' social similarities are present. A similar approach is used in [130] to deal with joint millimeter and micro wave dual base station resource allocation, in [131] for user base station association in small cell networks, and in [132] to optimize D2D offloading techniques. Caching in small cell networks can also be addressed as a many-to-many matching game [133]: by matching video popularity among users most frequently served by a given server it is possible to devise caching policies that minimize end-users' delays. Simulations show the approach is effective in small cell networks.

3) Social-Assisted Mobility Prediction: Motivated by the need to reduce the active scanning overhead in IEEE 802.11 networks, Wanalertlak *et al.* [40] propose a mobility prediction tool to anticipate the next access point a WiFi user is moving to. The proposed solution is based on context information on the handoffs which were performed in the past; specifically, the system stores centrally a time varying handoff table which is then fed into an ARIMA predictor which returns the likelihood of a given user to handoff to a specific access point. The quality of the predictor is measured in terms of signaling reduction due to active scanning.

The prediction of user mobility is also addressed in [134]. The authors leverage information coming from the social platform Foursquare to predict user mobility on coarse granularity. The *next check-in problem* is formulated to determine the next place in an urban environment which will be most likely visited by a user. The authors build a time-stamped dataset of "check-ins" performed by Foursquare users over a period of one month across several venues worldwide. A set of features is then defined to represent user mobility including user mobility features (e.g., number of historical visits to specific venues or categories of venues, number of historical visits that friends have done to specific venues), global mobility features (e.g., popularity of venues, distance between venues, transition frequency between couples of venues), and temporal features which measures the historical check-ins over specific time periods. Such a feature set is then used to train a supervised classification problem to predict the next check-in venue. Linear regression and M5 decision trees are used in this regard. The work is mostly speculative and does not address directly any specific application/use of the proposed mobility prediction tool.

Along the same lines, the mobility of users in urban environments is characterized in [135]. Different from the previous work which only exploits social information, the authors also leverage physical information about the current position of moving users. A probabilistic model of the mobile users' behavior is built and trained on a real life dataset of user mobility traces. A social-assisted mobility prediction model is proposed in [136], where a variable-order Markov model is developed and trained on both temporal features (i.e., when users were at specific locations) and social ones (i.e., when friends of specific users were at a given location). The accuracy of the proposed model is cross-validated on two user-mobility datasets.

4) Social-Assisted Radio Resource Allocation: The optimization of elastic traffic in the downlink of mobile radio networks is addressed in [137] and [138]. The key tenet is to provide to the downlink scheduler "richer" context to make better decisions in the allocation of the radio resources. Besides classical network-side context including the cell load

TABLE III
CONTEXT CLASSIFICATION SUMMARY: EACH CONTEXT IS ASSOCIATED TO ITS MOST POPULAR APPLICATIONS, PREDICTION TECHNIQUES, OPTIMIZATION METHODS AND MAIN NOTABLE CHARACTERISTICS

Context	Applications	Prediction[a]	Optimization	Remarks
Geographic [11-26, 28, 29, 31-35, 37, 38, 41-46, 131]	Mobility prediction Multimedia streaming Broadcast Resource allocation Duty cycling	1^{st} Probabilistic 2^{nd} Regression 3^{rd} Time series 4^{th} Classification	1) Prediction to define convex optimization problems 2) Prediction as the optimization objective	1) Prediction accuracy is inversely proportional to the time scale and granularity 2) High prediction accuracy can be obtained on long time scales if periodicity and/or trends are present 3) Prediction is more effectively used in delay tolerant applications
Link [30, 47-70, 72-79, 129, 158]	Channel forecast Resource allocation Network mapping Routing Multimedia streaming	1^{st} Regression 2^{nd} Time series 3^{rd} Probabilistic 4^{th} Classification	1) Markov decision process is used when statistical knowledge of the system is available 2) Convex optimization is preferred when it is possible to perform accurate forecast	1) Channel quality maps can be effectively used to improve networking 2) Mobility dynamics affect the prediction effectiveness 3) Channel is most often predicted by means of functional regression or Markovian models
Traffic [92-102, 104-120, 138, 145, 156, 165]	Traffic analysis Resource allocation Multimedia streaming	1^{st} Regression 2^{nd} Classification 3^{rd} Probabilistic	1) Maps are used to deterministically guide the optimization 2) Convex optimization problems can be formulated to obtain bounds	1) Improved long-term network optimization and reconfiguration 2) Traffic distribution is skewed both with regards to users and locations 3) Traffic has a strong time periodicity 4) Geo-localized information can be used as inputs for optimization
Social [40, 121-140, 148, 149, 154, 157, 159]	Network caching Mobility prediction Resource allocation Multimedia streaming	1^{st} Classification 2^{nd} Regression 3^{rd} Time series 4^{th} Probabilistic	1) Formal optimization problems can be defined, but they are usually impractical to be solved 2) Game theory and heuristics are the preferable online solutions	1) A fraction of social information can be accurately predicted 2) Prediction obtained from social information is usually coarse 3) Social information prediction can effectively improve application performance

[a]Ranking based on the number of papers reviewed in this survey using the predictor.

and the current channel quality indicator which are widely used in the literature to steer the scheduling, the authors propose to include user-side features which generically capture the satisfaction degree of the user for the reference application. Namely, the authors introduce the concept of a *transaction*, which represents the atomic data download requested by the end user (e.g., a Web page download via HTTP, an object download via HTTP or a file download via File Transfer Protocol (FTP)). For each transaction and for each application, a utility function is defined capturing the user's sensitivity with respect to the transmission delay and the expected completion time. The functional form of this utility function depends on the type of application which "generated" the transaction; as an example, the authors make the distinction between transactions from applications which are running in the foreground and the background on the user's terminal. For the sake of presentation, a parametric logistic function is used to represent the aforementioned utility. The authors then formulate an optimization problem to maximize the sum utility across all the users and transactions in a given mobile radio cell and design a greedy heuristic to obtain a sub-optimal solution in reasonable computing time. The proposed algorithm is validated against state-of-the-art scheduling solutions (PF / weighted PF scheduling) through simulation on synthetic data mimicking realistic user distributions, mobility patterns and traffic patterns.

In order to predict the spatial traffic of base stations in a cellular network, [139] applies the idea of social networks to base stations. Here, the base stations themselves create a social network and a social graph is created between them based on the spatial correlation of the traffic of each of them. The correlation is calculated using the Pearson coefficient. Based on the topology of the social graph, the most important base stations are identified and used for traffic prediction of the entire network, which is done using SVM. The authors conclude that with the traffic data of less than 10% of the base stations, effective prediction with less than 20% mean error can be achieved.

Social-oriented techniques related to the popularity of the end users are leveraged also in [140] where Tsiropoulos *et al.* target the performance optimization of downlink resource allocation in future generation networks. The utility maximization problem is formulated with the utility being a combination (product) of a network-oriented term (available bandwidth) and a social-oriented term (social distance). The social-oriented term is defined to be the degree centrality measure [141] for a specific user. The proposed problem is sub-optimally solved through a heuristic which is finally validated using synthetic data.

E. Summary

Hereafter, we summarize the main takeaways of the section in terms of application and objective for which different context types can be used. Table III provides a synthesis of the main considerations: each context is associated with its typical applications, prediction methodologies (ordered by decreasing popularity), optimization approaches and general remarks.

1) Mobility Prediction: It has been shown that predictability of user mobility can be potentially very high (93% potential predictability in user mobility as stated in [11]), despite the significant differences in the travel patterns. As a matter of fact, many papers study how to forecast users' mobility by means of a variety of techniques. For predicting trajectories, characterized by sequences of discretized locations indicated by cell identitys (IDs) or road segments, fixed-order Markov models or variable-order Markov models are the most promising tools, while for continuous trajectories, regression techniques are widely used. To enhance the prediction accuracy, the most popular ones leverage geographic information: GPS data, cell records and received signal strength are used to obtain precise and frequent data sampling to locate users on a map. However, the movements of an individual are largely influenced by those of other individuals via social relations. Several papers analyze social information and location check-ins to find recurrent patterns. For this second case usually a sparser dataset is available and may limit the accuracy of the prediction.

2) Network Efficiency: Predicting and optimizing network efficiency (i.e., increasing the performance of the network while using the same amount of resources) is the most frequent objective in anticipatory networking. We found papers exploiting all four types of context to achieve this. As such, objectives and constraints cover the whole attribute space. Improving network efficiency is likely to become the main driver for including anticipatory networking solutions in next generation networks.

3) Multimedia Streaming: The main source of data traffic in 4G networks has been multimedia streaming and, in particular, video on demand. 5G networks are expected to continue and even increase this trend. As a consequence, several anticipatory networking solutions focus on the optimization of this service. All the context types have been used to this extent and each has a different merit: social information is needed to predict when a given user is going to request a given content, combined geographic and social information allows the network to cache that content closer to where it will be required and physical channel information can be used to optimize the resource assignment.

4) Network Offloading: Mobility prediction can be used to handover communications between different technologies to decrease network congestion, improve user experience, reduce users' costs and increase energy efficiency.

5) Cognitive Networking: Physical channel prediction can be exploited for cognitive networking and for network mapping. The former application allows secondary users to access a shared medium when primary subscribers left resource unused, thus, predicting when this is going to happen will highly improve the effectiveness of the solution. The latter, instead, exploits link information to build networking maps that can provide other applications with an estimate of communication quality at a given time and place.

6) Throughput- and Traffic-Based Applications: Traffic information is usually studied to be, first, modeled and, subsequently, predicted. Traffic models and predictors are then used to improve networking efficiency by means of resource allocation, traffic shaping and network planning.

IV. Prediction Methodologies for Anticipatory Networking

In this section, we present some selected prediction methods for the types of context introduced in Section I-A. The selected methods are classified into four main categories: *time series methods*, *similarity-based classification*, *regression analysis*, and *statistical methods for probabilistic modeling*. Their mathematical principles and the application to inferring and predicting the aforementioned contextual information are introduced in Sections IV-A, IV-B, IV-C, and IV-D, respectively.

The goal of the prediction handbook is to show *which methods work in which situation*. In fact, selecting the appropriate prediction method requires to analyze the prediction variables and the model constraints with respect to the application scenario (see Section I-A). This section concludes with a series of takeaways that summarize some general principles for selection of prediction methods based on the scenario analysis.

A. Time Series Predictive Modeling

A time series is a set of time-stamped data entries which allows a natural association of data collected on a regular or irregular time basis. In wireless networks, large volumes of data are stored as time series and frequently show temporal correlation. For example, the trajectory of the mobile device can be characterized by successive time-stamped locations obtained from geographical measurements; individual social behavior can be expressed through time-evolving events; traffic loads modeled in time series can be leveraged for network planning and controlling. Fig. 3(a) and (b) illustrate two time series of per-cell and per-city aggregated uplink and downlink data traffic, where temporal correlation is clearly recognizable.

In the following, we introduce the two most widely used time series models based on linear dynamic systems: 1) AutoRegressive and Moving Average (ARMA), and 2) Kalman filters. Examples of context prediction in wireless networks are given and their extensions to nonlinear systems are briefly discussed.

1) Autoregressive and Moving Average Models: Consider a univariate time series $\{X_t : t \in \mathcal{T}\}$, where $\mathcal{T}$ denotes the set of time indices. The general ARMA model, denoted by ARMA(p, q), has p AR terms and q Moving Average (MA) terms, given by

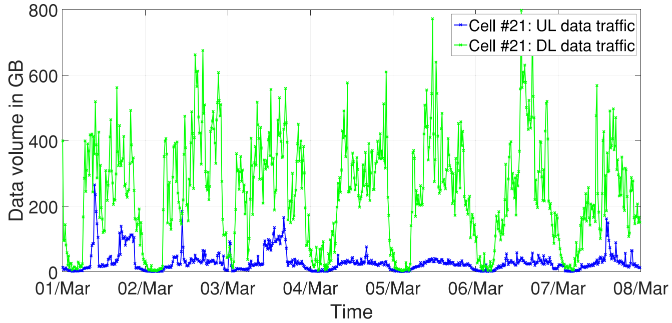
$$X_t = Z_t + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \theta_j Z_{t-j} \quad (1)$$

where Z_t is the process of the white noise errors, and $\{\phi_i\}_{i=1}^{p}$ and $\{\theta_j\}_{j=1}^{q}$ are the parameters. The ARMA model is a generalization of the simpler AR and MA models that can be obtained for $q = 0$ and $p = 0$ respectively. Using the *lag operator* $L^i X_t := X_{t-i}$ the model becomes

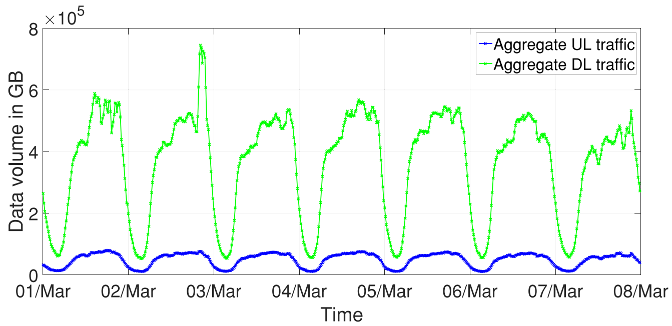
$$\phi(L) X_t = \theta(L) Z_t \quad (2)$$

where $\phi(L) := 1 - \sum_{i=1}^{p} \phi_i L^i$ and $\theta(L) := 1 + \sum_{j=1}^{q} \theta_j L^j$.

The fitting procedure of such processes assumes *stationarity*. However, this property is seldom verified in practice and *non-stationary* time series need to be stationarized through differencing and logging. The ARIMA model generalizes ARMA models for the case of non-stationary time series: a non seasonal ARIMA model ARIMA(p, d, q) after d differentiations



(a) Uplink and downlink traffic load in a cell grid in Rome, Italy.



(b) Aggregated uplink and downlink traffic load in Rome, Italy.

Fig. 3. Example of time series: Traffic load (aggregated every 15 minutes) for a week in March 2015 in Rome, Italy. Data source from Telecom Italia's Big Data Challenge [142].

reduces to an ARMA(p, q) of the form

$$\phi(L)\Delta^d X_t = \theta(L)Z_t, \tag{3}$$

where $\Delta^d = (1-L)^d$ denotes the dth difference operator.

Numerous studies have been done on prediction of traffic load in wireless or IP backbone networks using autoregressive models. The stationarity analysis often provides important clues for selecting the appropriate model. For instance, in [108] a low-order ARIMA model is applied to capture the non-stationary short memory process of traffic load, while in [109] a Gegenbauer ARMA model is used to specify long memory processes under the assumption of stationarity. Similar models are applied to mobility- or channel-related contexts. In [40], an exponential weighted moving average, equivalent to ARIMA$(0, 1, 1)$, is used to forecast handoffs. In [13] and [47], AR models are applied to predict future signal-to-noise ratio values and user positions, respectively. If the variance of the data varies with time, as in [110] for data traffic, and can be expressed using an ARMA, then the whole model is referred to as GARCH.

2) Kalman Filter: Kalman filters are widely applied in time series analysis for linear dynamic systems, which track the estimated system state and its uncertainty variance. In the anticipatory networking literature, Kalman filters have been mainly adopted to model the linear dependence of the system states based on historical data.

Consider a multivariate time series $\{\mathbf{x}_t \in \mathbb{R}^n : t \in \mathcal{T}\}$, the Kalman filter addresses the problem of estimating state $\mathbf{x}_t$ that is governed by the linear stochastic difference equation

$$\mathbf{x}_t = \mathbf{A}_t\mathbf{x}_{t-1} + \mathbf{B}_t\mathbf{u}_t + \mathbf{w}_t, \;\; t = 0, 1, \ldots, \tag{4}$$

where $\mathbf{A}_t \in \mathbb{R}^{n\times n}$ expresses the state transition, and $\mathbf{B}_t \in \mathbb{R}^{n\times l}$ relates the optional control input $\mathbf{u}_t \in \mathbb{R}^l$ to the state $\mathbf{x}_t \in \mathbb{R}^n$. The random variable $\mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t)$ represents a multivariate normal noise process with covariance matrix $\mathbf{Q}_t \in \mathbb{R}^{n\times n}$. The observation $\mathbf{z}_t \in \mathbb{R}^m$ of the true state $\mathbf{x}_t$ is given by

$$\mathbf{z}_t = \mathbf{H}_t\mathbf{x}_t + \mathbf{v}_t, \tag{5}$$

where $\mathbf{H}_t \in \mathbb{R}^{m\times n}$ maps the true state space into the observed space. The random variable $\mathbf{v}_t$ is the observation noise process following $\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_t)$ with covariance $\mathbf{R}_t \in \mathbb{R}^{n\times n}$. Kalman filters iterate between 1) predicting the system state with Eq. (4) and 2) updating the model according to Eq. (5) to refine the previous prediction. The interested reader is referred to [143] for more details.

In [32] and [144], Kalman filters are used to study users' mobility. Wireless channel gains are studied in [49] with KKF, while Okutani and Stephanedes [145] adopt the technique to predict short-term traffic volume. The extended Kalman filter adapts the standard model to nonlinear systems via online Taylor expansion. According to [146], this improves shadow/fading estimation.

B. Similarity-Based Classification

Similarity-based classification aims to find inherent structures within a dataset. The core rationale is that similarity patterns in a dataset can be used to predict unknown data or missing features. Recommendation systems are a typical application where users give a score to items and the system tries to infer similarities among users and scores to predict the missing entries.

These techniques are unsupervised learning methods, since categories are not predetermined, but are inferred from the data. They are applied to datasets exhibiting one or more of the following properties: 1) entries of the dataset have many attributes, 2) no law is known to link the different features, and 3) no classification is available to manually label the dataset.

In what follows, we briefly review the similarity-based classification tools that have been used in the anticipatory networking literature accounted for in this survey.

1) Collaborative Filtering: Recommendation systems usually adopt Collaborative Filtering (CF) to predict unknown opinions according to user's and/or content's similarities. While a thorough survey is available in [147], here, we just introduce the main concepts related to anticipatory networking.

CF predicts the missing entries of a $n_c \times n_u$ matrix $\mathbf{Y} \in \mathcal{A}^{n_c\times n_u}$, mapping n_c users to n_u contents through their opinions which are taken from an alphabet $\mathcal{A}$ of possible ratings. Thus, the entry $y_{ik}, i \in \{1, \ldots, n_c\}, k \in \{1, \ldots, n_u\}$ expresses how much user k likes content i. An auxiliary matrix $\mathbf{R} \in [0, 1]^{n_c\times n_u}$ expresses whether user k evaluated content i ($r_{ik} = 1$) or not ($r_{ik} = 0$).

To predict the missing entries of $\mathbf{Y}$ the feature learning approach exploits a set of n_f features to represent contents' and users' similarities and defines two matrices $\mathbf{X} \in [0, 1]^{n_c\times n_f}$ and $\mathbf{\Theta} \in \mathcal{A}^{n_u\times n_f}$, whose entries x_{ij} and θ_{kj} represent how much content i is represented by feature j and how high user k would rate a content completely defined by feature j, respectively. The new matrices aim to map $\mathbf{Y}$ in the feature space and they can be computed by:

$$\underset{\mathbf{X},\boldsymbol{\Theta}}{\operatorname{argmin}} \sum_{i,k:r_{ik}=1} \left(\mathbf{x}_{i*}\boldsymbol{\theta}_{k*}^T - y_{ik}\right)^2, \tag{6}$$

where $\mathbf{x}_{i*} := (\mathrm{col}_i\mathbf{X}^T)^T$ denotes the i-th row of matrix $\mathbf{X}$. Note that in (6) the regularization terms are omitted. Solving (6) amounts to obtain a matrix $\tilde{\mathbf{Y}} = \mathbf{X}\boldsymbol{\Theta}^T$ which best approximates $\mathbf{Y}$ according to the available information ($i, k : r_{ik} = 1$). Finally, $\tilde{y}_{ik} = \mathbf{x}_{i*}\boldsymbol{\theta}_{k*}^T$ predicts how user k with parameters $\boldsymbol{\theta}_{k*}$ rates content i having feature vector $\mathbf{x}_{i*}$.

Other applications of CF are, for instance, network caching optimization [148], [149], where communication efficiency is optimized by storing contents where and when they are predicted to be consumed. Similarly, location-based services [134] predict where and what to serve to a given user.

2) Clustering: Clustering techniques are meant to group elements that share similar characteristics. The following provides an introduction to K-means, which is among the most commonly-used clustering techniques in anticipatory networking. The interested reader is referred to [150] for a complete review.

K-means splits a given dataset into K groups without any prior information about the group structure. The basic idea is to associate each observation point from a dataset $\mathcal{X} := \{\mathbf{x}_i \in \mathbb{R}^n : i = 1, \dots, M\}$, to one of the centroids in set $\mathcal{M} := \{\boldsymbol{\mu}_j \in \mathbb{R}^n : j = 1, \dots, K\}$. The centroids are optimized by minimizing the intra-cluster sum of squares (sum of distance of each point in the cluster to the K centroids), given by

$$\underset{\mathcal{C},\mathcal{M}}{\text{minimize}} \sum_{j=1}^{K}\sum_{i=1}^{M} c_{ij}\left\|\mathbf{x}_i - \boldsymbol{\mu}_j\right\|^2, \tag{7}$$

where $\mathcal{C} := \{c_{ij} \in \{0, 1\} : i = 1, \dots, M, j = 1, \dots, K\}$ associates entry $\mathbf{x}_i$ to centroid $\boldsymbol{\mu}_j$. No entry can be associated to multiple centroids ($\sum_{j=1}^{K} c_{ij} = 1, \forall i \in \mathcal{M}$).

Clustering is applied in anticipatory networking to build a data-driven link model [51], to find similarities within vehicular paths [34], to identify social events [99] that might impact network performance, and to identify device types [93].

3) Decision Trees: A supervised version of clustering is *decision tree learning* (the interested reader is referred to [151] for a survey on the topic). Assuming that each input observation is mapped to a consequence on its target value (such as reward, utility, cost, etc.), the goal of decision tree learning is to build a set of rules to map the observations to their target values. Each decision branches the tree into different paths that lead to leaves representing the class labels. With prior knowledge, decision trees can be exploited for location-based services [134], for identifying trajectory similarities [35], and for predicting the QoE for multimedia streams [101]. For continuous target variables, regression trees can be used to learn trends in network performance [98].

C. Regression Analysis

When the interest lies in understanding the relationship between different variables, regression analysis is used to predict dependent variables from a number of independent variables by means of so-called regression functions. In the following, we introduce three regression techniques, which are able to capture complex nonlinear relationships, namely *functional regression*, *support vector machines* and *artificial neural networks*.

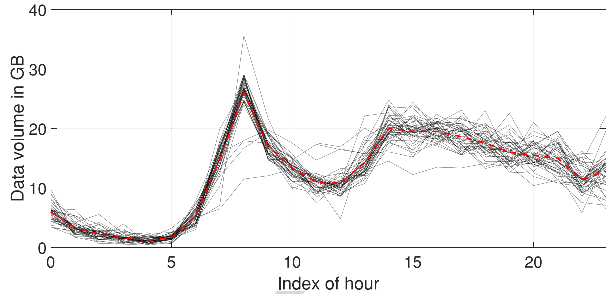


Fig. 4. Example of a functional dataset: WiFi traffic in Rome depending on hour of the day. Data source from Telecom Italia's Big Data Challenge [142].

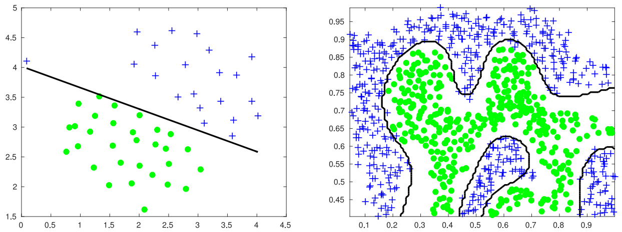
1) Functional Regression: Functional data often arise from measurements, where each point is expressed as a function over a physical continuum (e.g., Fig. 4 illustrates the example of aggregated WiFi traffic as a function of the hour of the day). Functional regression has two interesting properties: smoothness allows to study derivatives, which may reveal important aspects of the processes generating the data, and the mapping between original data and the functional space may reduce the dimensionality of the problem and, as a consequence, the computational complexity [152]. The commonly encountered form of function prediction regression model (scalar-on-function) is given by [153]:

$$Y_i = B_0 + \int X_i(z)B(z)dz + E_i \tag{8}$$

where $Y_i, i = 1, \dots, M$ is a continuous response, $X_i(z)$ is a functional predictor over the variable z, $B(z)$ is the functional coefficient, B_0 is the intercept, and E_i is the residual error.

Functional regression methods are applied in [94] to predict traffic-related Long Term Evolution (LTE) metrics (e.g., throughput, modulation and coding scheme, and used resources) showing that cloud analytics of short-term LTE metrics is feasible. In [154], functional regression is used to study churn rate of mobile subscribers to maximize the carrier profitability.

2) Support Vector Machines: SVM is a supervised learning technique that constructs a hyperplane or set of hyperplanes (linear or nonlinear) in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks. In this survey we introduce the SVM for classification, and the same principle is used by SVM for regression. Consider a training dataset $\{(\mathbf{x}_i, y_i) : \mathbf{x}_i \in \mathbb{R}^n, y_i \in \{-1, 1\}, i = 1, \dots, M\}$, where $\mathbf{x}_i$ is the i-th training vector and y_i the label of its class. First, let us assume that the data is linearly separable and define the linear separating hyperplane as $\mathbf{w} \cdot \mathbf{x} - b = 0$, where $\mathbf{w} \cdot \mathbf{x}$ is the Euclidean inner product. The optimal hyperplane is the one that maximizes the *margin* (i.e., distance from the hyperplane to the instances closest to it on either side), which



(a) Linear (b) Gaussian

Fig. 5. Examples of SVM, where different datasets are analyzed according to a linear (left) and a Gaussian (right) kernel.

can be found by solving the following optimization problem:

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}||\mathbf{w}||^2 \\ &\text{subject to} \quad y_i(\mathbf{x}_i \cdot \mathbf{w} + b) - 1 \geq 0 \ \forall i \in \{1, \dots, M\}. \end{aligned} \tag{9}$$

Fig. 5(a) shows an example of linear SVM classifier separating two classes in $\mathbb{R}^2$.

If the data is not linearly separable, the training points are projected to a high-dimensional space $\mathcal{H}$ through a nonlinear transformation $\boldsymbol{\phi} : R^n \rightarrow \mathcal{H}$. Then, a linear model in the new space is built, which corresponds to a nonlinear model in the original space. Since the solution of (9) consists of inner products of training data $\mathbf{x}_i \cdot \mathbf{x}_j$, for all i, j, in the new space the solution is in the form of $\boldsymbol{\phi}(\mathbf{x}_i) \cdot \boldsymbol{\phi}(\mathbf{x}_j)$. The *kernel trick* is applied to replace the inner product of basis functions by a *kernel function* $K(\mathbf{x}_i, \mathbf{x}_j) = \boldsymbol{\phi}(\mathbf{x}_i) \cdot \boldsymbol{\phi}(\mathbf{x}_j)$ between instances in the original input space, without explicitly building the transformation $\boldsymbol{\phi}$.

The Gaussian kernel $K(\mathbf{x}, \mathbf{y}) := \exp(\gamma||\mathbf{x} - \mathbf{y}||^2)$ is one of the most widely used kernels in the literature. For example, it is used in [15] to predict user mobility. Kasparick *et al.* [52] propose an algorithm for reconstructing coverage maps from path-loss measurements using a kernel method. Nevertheless, choosing an appropriate kernel for a given prediction task remains one of the main challenges.

3) Artificial Neural Networks: ANN is a supervised machine learning solution for both regression and classification. An ANN is a network of nodes, or *neurons*, grouped into three layers (input, hidden and output), which allows for nonlinear classification. Ideally, it can achieve zero training error.

Consider a training dataset $\{(\mathbf{x}_i, y_i) : \mathbf{x}_i \in \mathbb{R}^n, i = 1, \dots, M\}$. Each hidden node h_l approximates a so-called logistic function in the form $h_l = 1/(1 + \exp(-\boldsymbol{\omega}_l \cdot \mathbf{x}))$, where $\boldsymbol{\omega}_l$ is a weight vector. The outputs of the hidden nodes are processed by the output nodes to approximate $\mathbf{y}$. These nodes use linear and logistic functions for regression and classification, respectively. In the linear case, the approximated output is represented as:

$$\hat{\mathbf{y}} = \sum_{l=1}^{L} h_l v_l = \sum_{l=1}^{L} \frac{1}{1 + \exp(-\boldsymbol{\omega}_l \cdot \mathbf{x})} v_l, \tag{10}$$

where L is the number of hidden nodes and v_l is the weight vector of the output layer. The training of an ANN can be performed by means of the *backpropagation* method that finds weights for both layers to minimize the mean squared error between the training labels y and their approximations $\hat{y}$. In the anticipatory networking literature, ANNs have been used for example to predict mobility in mobile ad-hoc networks [14], [155].

For both SVMs and ANNs, as for other supervised learning approaches, no prior knowledge about the system is required but a large training set has to be acquired for parameter setting in the predictive model. A careful analysis needs to be performed while processing the training data in order to avoid both overfitting and underlearning.

D. Statistical Methods for Probabilistic Forecasting

Probabilistic forecasting involves the use of information at hand to make statements about the likely course of future events. In the following subsections, we introduce two probabilistic forecasting techniques: *Markovian models* and *Bayesian inference*.

1) Markovian Models: These models can be applied to any system for which state transitions only depend on the current state. In the following we briefly discuss the basic concepts of discrete, and continuous time Markov Chains (MCs) and their respective applications to anticipatory networking.

A Discrete Time Markov Chain (DTMC) is a discrete time stochastic process $X_n (n \in \mathbb{N})$, where a state X_n takes a finite number of values from a set $\mathcal{X}$ in each time slot. The Markovian property for a DTMC transitioning from any time slot k to $k+1$ is expressed as follows:

$$P(X_{k+1} = j | X_k = i) = p_{ij}(k). \tag{11}$$

For a stationary DTMC, the subscript k is omitted and the transition matrix $\mathbf{P}$, where p_{ij} represents the transition probability from state i to state j, completely describes the model. Empirical measurements on mobility and traffic evolution can be accurately predicted using a DTMC with low computational complexity [19], [23], [26], [93], [136]. However, obtaining the transition probabilities of the system requires a variable training period, which depends on the prediction goal. In practice, the data collection period can be in the order of one [93] or even multiple weeks [20], [53].

A DTMC assumes the time the system spends in each state is equal for all states. This time depends on the prediction application and can range from a few hundred milliseconds to predict wireless channel quality [62], to tens of seconds for user mobility prediction [19], [53], to hours for Internet traffic [93]. For tractability reason, the state space is often compressed by means of simple heuristics [20], [53], [102], K-means clustering [62], [136], equal probability classification [102], and density-based clustering [136].

Eq. (11) defines a first order DTMC and can be extended to the l-th order (i.e., transition probabilities depend on the l previous states). By Using higher order, DTMCs can increase the accuracy of the prediction at the expense of a longer training time and an increased computational complexity [19], [23], [136].

If the sojourn time of each state is relevant to the prediction, the system can be modeled as a Continuous Time Markov Chain (CTMC). The Markovian property is preserved in

CTMC when the sojourn time is exponentially distributed, as in [21]. When the sojourn time has an arbitrary distribution, it becomes a Markov renewal process as described in [17] and [18].

If the transition probabilities cannot be directly measured, but only the output of the system is quantifiable (dependent on the state), hidden Markov models allow to map the output state space to the unobservable model that governs the system. As an example, the inter-download times of video segments are predicted in [102], where the output sequences are the inter-download times of the already downloaded segments and the states are the instants of the next download request.

2) Bayesian Inference: This approach allows to make statements about what is unknown, by conditioning on what is known. Bayesian prediction can be summarized in the following steps: 1) define a *model* that expresses qualitative aspects of our knowledge but has unknown parameters, 2) specify a *prior* probability distribution for the unknown parameters, 3) compute the *posterior* probability distribution for the parameters, given the observed data, and 4) make predictions by averaging over the posterior distribution.

Given a set of observed data $\mathcal{D} := \{(\mathbf{x}_i, \mathbf{y}_i) : i = 1, \ldots, M\}$ consisting of a set of input samples $\mathcal{X} := \{\mathbf{x}_i \in \mathbb{R}^p : i = 1, \ldots, M\}$ and a set of output samples $\mathcal{Y} := \{\mathbf{y}_i \in \mathbb{R}^q : i = 1, \ldots, M\}$, inference in Bayesian models is based on the *posterior distribution* over the parameters, given by the *Bayes' rule*:

$$p(\boldsymbol{\theta}|\mathcal{D}) = \frac{p(\mathcal{Y}|\mathcal{X}, \boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathcal{Y}|\mathcal{X})} \propto p(\mathcal{Y}|\mathcal{X}, \boldsymbol{\theta})p(\boldsymbol{\theta}), \tag{12}$$

where $\boldsymbol{\theta}$ is the unknown parameter vector.

Two recent works adopting the Bayesian framework are [38] and [55]. The former focuses on spatial prediction of the wireless channel, building a 2D non-stationary random field accounting for pathloss, shadowing and multipath. The latter exploits spatial and temporal correlation to develop a general prediction model for the channel gain of mobile users.

E. Summary

Hereafter, we provide some guidelines for selecting the appropriate prediction methods depending on the application scenario or context of interest.

1) Applications and Data: The predicted context is the most important information that drives decision making in anticipatory optimization problems (see Section V). Thus, the selection of the prediction method shall take into consideration the objectives of the application and the constraints imposed by the available data.

a) Choosing the outputs: Applications define the properties of the predicted variables, such as dimension, granularity, accuracy, and range. For example, large granularity or high data aggregation (such as frequently visited location, social behavior pattern) is best dealt with similarity-based classification methods which provide sufficiently accurate prediction without the complexity of other model-based regression techniques.

b) System model and data: The application environment is equally important as its outputs, which determines the constraints of modeling. Often, an accurate analysis of the scenario might highlight linearity, deterministic and/or causal laws among the variables that can further improve the prediction accuracy. Moreover, the quality of dataset heavily affects the prediction accuracy. Different methods exhibit different level of robustness to noisy data.

2) Guidelines for Selecting Methods: To choose the correct tool among the aforementioned set, we study the rationale for adopting each of them in the literature and derive the following practical guidelines.

a) Model-based methods: When a physical model exists, model-based regression techniques based on closed-form expressions can be used to obtain an accurate prediction. They are usually preferable for long-term forecast and exhibit good resilience to poor data quality.

b) Time series-based methods: These are the most convenient tools when the information is abundant and shows strong temporal correlation. Under these conditions, time series methods provide simple means to obtain multiple scale prediction of moderate to high precision.

c) Causal methods: If the data exhibits large and fast variations, causality laws can be key to obtain robust predictions. In particular, if a causal relationship can be observed between the variables of interest and the other observable data, causal models usually outperform pure data-driven models.

d) Probabilistic models: If the physical model of the prediction variable is either unavailable or too complex to be used, probabilistic models offer robust prediction based on the observation of a sufficient amount of data. In addition, probabilistic methods are capable of quantifying the uncertainty of the prediction, based on the probability density function of the predicted state.

3) Prediction Summary: Table IV characterizes each prediction method with respect to *properties of the context* and *constraints* presented in Section I-A. Note that the methods for predicting a multivariate process can be applied to univariate processes without loss of generality. The granularity of variables and the prediction range are described using qualitative attributes such as **S**hort, **M**edium, **L**arge, and **any** instead of explicit values. For example, for the time series of traffic load per cell, S, M and L time scales are generally defined by minutes, tens of minutes and hours, respectively, while for the time series of channel gain, they can be seen as milliseconds, hundreds of milliseconds and seconds, respectively. The sixth column reports the prediction type, that can be driven by **data**, **models** or **both**. Linearity indicates whether it is required (**Y**) or not (**N**) or applicable in **both** cases. The side information column states whether out-of-band information can (**both**), cannot (**N**) or must (**Y**) be used to build the model. Finally, the quality column reports whether the predictor is **weak** or **robust** against insufficient or unreliable dataset.

V. Optimization Techniques for Anticipatory Networking

This section identifies the main optimization techniques adopted by anticipatory networking solutions to achieve their

TABLE IV
SELECTED PREDICTION METHODS: VARIABLES OF INTEREST AND CONSTRAINTS OF MODELING

Prediction Method		Properties of the Context			Constraints			
Class	Methodology	Dimension	Granularity	Range	Type	Linearity	Side Info.	Quality
Time series	ARIMA	univariate	M/L	S	data	Y	N	weak
	Kalman filter	multivariate	M/L	S	data	Y	N	weak
	References	*ARIMA*: [13], [38], [40], [46], [47], [54], [58], [59], [63], [100], [119] *Kalman*: [32], [49]						
Classification	CF	multivariate	L	M/L	data	Y	both	robust
	Clustering	multivariate	L	M/L	data	both	both	robust
	Decision trees	multivariate	L	any	data	both	Y	robust
	References	*CF*: [16], [134], [149] *Cluster*: [15], [34], [51], [117], [122], [123], [148], [156] *Decision trees*: [35], [98], [101]						
Regression	Functional	multivariate	any	M/L	models	both	Y	robust
	SVM	multivariate	any	any	both	both	both	weak
	ANN	multivariate	any	any	data	both	both	weak
	References	*Functional*: [28], [29], [38], [64], [99], [104], [105] *SVM*: [51], [114], [139] *ANN*: [14], [48], [106], [107]						
Probabilistic	Markovian	multivariate	M/L	any	both	both	both	weak
	Bayesian	multivariate	any	any	both	both	Y	weak
	References	*Probabilistic*: [12], [16]–[21], [23]–[26], [30], [50], [53], [60], [61], [93], [102], [116], [136], [157] *Bayesian*: [33], [37], [58], [126], [127], [129], [130], [132], [135], [158], [159]						

AQ5

TABLE V
OPTIMIZATION METHODS SUMMARY

Methodology	Properties of context	Modeling constraints
ConvOpt	Can support any context property, but larger system states slow the solver performance. The solution accuracy is linked to the context precision.	Linearity can be exploited to improve the solver efficiency, while data reliability impacts the solution optimality.
MPC	Usually offers the highest precision by coupling prediction and optimization.	The most computationally intensive technique.
MDP	Limited range and precision.	The most robust approach to low data reliability. Although the system setup can be computationally intensive, it allows for lightweight policies to be implemented.
Game theory	Limited granularity to allow the system to converge to an equilibrium.	Very low computational complexity. Fast dynamics hinder the system convergence.

objectives. Disregarding the particular domain of each work, the common denominator is to leverage some future knowledge obtained by means of prediction to drive the system optimization. How this optimization is performed depends both on the ultimate objectives and how data are predicted and stored.

In general, we found two main strategies for optimization: (1) adopting a well-known optimization framework to model the problem and (2) designing a novel solution (most often) based on heuristic considerations about the problem. The two strategies are not mutually exclusive and often, when known approaches lead to too complex or impractical solutions, they are mixed in order to provide feasible approximation of the original problem.

Heuristic approaches usually consist of (1) algorithms that allow for fast computation of an approximation of the solution of a more complex problem (e.g., convex optimization) and (2) greedy approaches that can be proven optimal under some set of assumptions. Both approaches trade optimality for complexity and most often are able to obtain performance quite close to the optimal one. However, heuristic approaches are tailored to the specific application and are usually difficult to be generalized or to be adapted for different scenarios, thus they cannot be directly applied to new applications if the new requirements do not match those of the original scenario.

In what follows, we focus on optimization methods only and we will provide some introductory descriptions of the most relevant ones used for anticipatory networking. The objective is to provide the reader with a minimum set of tools to understand the methodologies and to highlight the main properties and applications.

A. Convex Optimization

Convex optimization is a field that studies the problem of minimizing a convex function over convex sets. The interested reader can refer to [160] for convex optimization theory and algorithms. Hereafter, we will adopt Boyd's notation [160] to introduce definitions and formulations that frequently appear in anticipatory networking papers.

The inputs are often referred to as the optimization variables of the problem and defined as the vector $\mathbf{x} = (x_1, \ldots, x_n)$. In order to compute the best configuration or, more precisely, to optimize the variables, an objective is defined: this usually corresponds to minimizing a function of the optimization variables, $f_0 : \mathbb{R}^n \to \mathbb{R}$. The feasible set of input configurations is usually defined through a set of m constraints $f_i(x) \leq b_i$, $i = 1, \ldots, m$, with $f_i : \mathbb{R}^n \to \mathbb{R}$. The general formulation of the problem is

$$\begin{aligned} &\text{minimize} \;\; f_0(\mathbf{x}) \\ &\text{subject to} \;\; f_i \leq b_i, \quad i = 1, \ldots, m. \end{aligned} \tag{13}$$

The solution to the optimization problem is an optimal vector $\mathbf{x}^*$ that provides the smallest value of the objective function, while satisfying all the constraints.

The convexity property (i.e., objective and constraint functions satisfy $f_i(a\mathbf{x} + (1-a)\mathbf{y}) \leq af_i(\mathbf{x}) + (1-a)f_i(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $a \in [0, 1]$) can be exploited in order to derive efficient algorithms that allows for fast computation of the optimal solution. Furthermore, if the optimization function and the constraints are linear, i.e., $f_i(a\mathbf{x}+b\mathbf{y}) = af_i(\mathbf{x})+bf_i(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $a, b \in \mathbb{R}$, the problem belongs to the class of *linear optimization*. For this class, highly efficient solvers exist, thanks to their inherently simple structure. Within the linear optimization class, three subclasses are of particular interest for anticipatory networking: least-squares problems, linear programs and mixed-integer linear programs.

Least-squares problems can be thought of as distance minimization problems. They have no constraints ($m = 0$) and their general formulation is:

$$\text{minimize } f_0(\mathbf{x}) = ||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2, \tag{14}$$

where $A \in \mathbb{R}^{k\times n}$, with $k \geq n$ and $||x||_2$ is the Euclidean norm. Notably, problems of this class have an analytical solution $\mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$ (where superscript ^{T} denotes the transpose) derived from reducing the problem to the set of linear equations $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$.

Linear programming (LP) problems are characterized by linear objective function and constraints and are written as

$$\begin{aligned} &\text{minimize } && \mathbf{c}^T\mathbf{x} \\ &\text{subject to } && \mathbf{A}^T\mathbf{x} \leq b, \end{aligned} \tag{15}$$

where $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{n\times m}$ and $\mathbf{b} \in \mathbb{R}^n$ are the parameters of the problem. Although, there is no analytical closed-form solution to LP problems, a variety of efficient algorithms are available to compute the optimal vector $\mathbf{x}^*$. When the optimization variable is a vector of integers $x \in \mathbb{Z}^n$, the class of problems is called *integer linear programming* (ILP), while the class of *mixed-integers linear programming* (MILP) allows for both integer and real variables to co-exist. These last two classes of problems can be shown to be NP-hard (while LP is P complete) and their solution often implies combinatorial aspects. See [161] for more details on integer optimization.

In anticipatory networking, we find that resource allocation problems are often modeled as LP, ILP or MILP, by setting the amount of resources to be allocated as the optimization variable and accounting for prediction in the constraints of the problem. In [72], prediction of the channel gain is exploited to optimize the energy efficiency of the network. Time is modeled as a finite number of slots corresponding to the look-ahead time of the prediction. When dealing with multimedia streaming, the data buffer is usually modeled in the constraints of the problem by linking the state at a given time slot to the previous slot. The solver will then choose whether to use resources in the current slot or use what has been accumulated in the buffer, as in, e.g., [77]. Admission control is often used to enforce quality-of-service, e.g., [74] and [156], with the drawback of introducing integer variables in the optimization function. In these cases, the optimal ILP/MILP formulation is followed by a fast heuristic that enables the implementation of real-time algorithms.

B. Model Predictive Control

Model Predictive Control (MPC) is a control theoretic approach that optimizes the sequence of actions in a dynamic system by using the process model of that system within a finite time horizon. Therefore, the process model, i.e., the process that turns the system from one state to the next, should be known. In each time slot t, the system state, $\mathbf{x}(t)$, is defined as a vector of attributes that define the relevant properties of the system. At each state, the control action, $\mathbf{u}(t)$, turns the system to the next state $\mathbf{x}(t+1)$ and results in the output $\mathbf{y}(t+1)$. In case the system is linear, both the next state and the output can be determined as follows:

$$\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) + \boldsymbol{\psi}(t) \tag{16}$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \boldsymbol{\epsilon}(t), \tag{17}$$

where $\boldsymbol{\psi}(t)$ and $\boldsymbol{\epsilon}(t)$ are usually zero mean random variables used to model the effect of disturbances on the input and output, respectively, and $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are matrices determined by the system model.

At each time slot, the next N states and their respective outputs are predicted and a cost function $J(\cdot)$ is minimized to determine the optimal control action $\mathbf{u}^*(t)$ at $t = t_0$:

$$\mathbf{u}^*(t_0) = \arg\min_{\mathbf{u}(t_0)} J\big(\hat{\mathbf{x}}(t_0), \mathbf{u}(t_0)\big), \tag{18}$$

where $\hat{\mathbf{x}}(t_0)$ is the set of all the predicted states from $t = t_0+1$ to $t = t_0 + N$, including the observed state at $t = t_0$. The expression in (18) essentially states that the optimal action of the current time slot is computed based on the predicted states of a finite time horizon in the future. In other words, in each time slot the MPC sequentially performs a N step lookahead open loop optimization of which only the first step is implemented [162].

This approach has been adopted for on-line prediction and optimization of wireless networks [100], [158]. Since the process model (for the prediction of future states and outputs) is available in this kind of systems, autoregressive methods can be used along with Kalman filtering [100], or max-min MPC formulation [159]. In [158], Kalman filtering is compared to other methods such as mean and median value estimation, Markov chains, and exponential averaging filters.

Optimization based on MPC relies on a finite horizon. The length of the horizon determines the trade-off between complexity and accuracy. Longer horizons need further look ahead and more complex prediction but in turn result in a more foresighted control action [159]. Reducing the horizon reduces the complexity while resulting in a more myopic action. This trade-off is examined in [158] by proposing an algorithm that adaptively adjusts the horizon length. In general, the prediction horizon is kept to a fairly low number (1 step in [159] and 6 steps in [100]) to avoid high computation overhead.

It is worth noting that MPC methods can be extended to the nonlinear case. In this case, the prediction accuracy and control optimality increase at the cost of more complex algorithms to find the solution [162]. Another benefit of these approaches is their applicability to non-stationary problems.

C. Markov Decision Process

Markov Decision Process (MDP) is an efficient tool for optimizing sequential decision making in stochastic environments. Unlike MPCs, MDPs can only be applied to stationary systems where a priori information about the dynamics of the system as well as the state-action space is available.

A MDP consists of a four tuple $(\mathcal{X}, \mathcal{U}, \mathbf{P}, r)$, where $\mathcal{X}$ and $\mathcal{U}$ represent the set of all achievable states in the system and the set of all actions that can be performed in each of the states, respectively. Time is assumed to be slotted and in any time slot t, the system is in state $x_t \in \mathcal{X}$ from which it can take an action u_t from the set $U_{x_t} \in \mathcal{U}$. Due to the assumption of stationarity, we can omit the time subscript for states and actions. Upon taking action u in state x, the system moves to the next state $x' \in \mathcal{X}$ with transition probability $\mathbf{P}(x'|x, u)$ and receives a reward equal to $r(x, u, x')$. The transition probabilities are predicted and modeled as a Markov Chain prior to solving the MDP and preserve the Markovian behavior of the system.

The goal is to find the optimal policy $\pi^* : \mathcal{X} \rightarrow \mathcal{U}$ (i.e., optimal sequence of actions that must be taken from any initial state) in order to maximize the long term discounted average reward $\mathbb{E}\left(\sum_{t=0}^{\infty} \gamma^t r(x_t, u_t, x_{t+1})\right)$, where $0 \leq \gamma < 1$ is called *discount factor* and determines how myopic (if closer to zero) or foresighted (if closer to 1) the decision process should be. In order to derive the optimal policy, each state is assigned to a value function $V^{\pi}(x)$, which is defined as the long term discounted sum of rewards obtained by following policy π from state x onwards. The goal of MDP algorithms is to find $V^{\pi^*}(x) (\forall x \in \mathcal{X})$. Given that the Markovian property holds, it has been proved that the optimal value functions follow the Bellman optimality criterion described below [163]:

$$V^{\pi^*}(x) = \max_{u \in \mathcal{U}} \sum_{x' \in \mathcal{X}'} \Big(r\big(x, u, x'\big) + \gamma \mathbf{P}\big(x'|x, u\big) V^{\pi^*}(x') \Big) \quad \forall x \in \mathcal{X}, \tag{19}$$

where $\mathcal{X}' \subset \mathcal{X}$ is the set of states for which $\mathbf{P}(x'|x, u) > 0$. In order to solve the above equation set, linear programming or dynamic programming techniques can be used, in which the optimal policy is derived by simple iterative algorithms such as policy iteration and value iteration [163].

MDPs are very efficient for several problems, especially in the framework of anticipatory networking, due to their wide applicability and ease of implementation. MDP-based optimized download policies for adaptive video transmission under varying channel and network conditions are presented in [60], [62], and [157].

In order to avoid large state spaces (which limit the applicability of MDPs), there are cases where the accuracy of the model must be compromised for simplicity. In [157], a large video receiver buffer is modeled for storing video on demand but only a small portion of the buffer is used in the optimization, while the rest of the buffer follows a heuristic download policy. References [60] and [62] solve this problem by increasing the duration of the time slot such that more video can be downloaded in each slot and, therefore, the buffer is filled entirely based on the optimal policy. This, in turn, comes at the cost of lower accuracy, since the assumption is that the system is static within the duration of a time slot. Heuristic approaches are also adopted for on-line applications. For instance, creating decision trees with low depth from the MDP outputs is proposed in [62]. Simpler heuristics are also applied to the MDP outputs in [60], [149], and [157].

If any of the assumptions discussed above does not hold, or if the state space of the system is too large, MDPs and their respective dynamic programming solution algorithms fail. However, there are alternative techniques to solve this kind of problems. For instance, if the system dynamics follow a Markov Renewal Process instead of a MC, a semi MDP is solved instead of the regular one [163]. In non-stationary systems, for which the dynamics cannot be predicted a priori or the reward function is not known beforehand, reinforcement learning [164] can be applied and the optimization turns into an on-line unsupervised learning problem. Large state spaces can be dealt with using value function approximation, where the value function of the MDP is approximated as a linear function, a neural network, or a decision tree [164]. If different subsets of state attributes have independent effects on the overall reward, i.e., multi user resource allocation, the problem can be modeled as a weakly coupled MDP [165] and can be decomposed into smaller and more tractable MDPs.

D. Game Theoretic Approaches

Although small in number, the papers adopting a game theoretic framework offer an alternative approach to optimization. In fact, while the approaches described in the previous subsections strive to compute the optimal solution of an often complex problem formulation, game theory defines policies that allow the system to converge towards a so-called equilibrium, where no player can modify her action to improve her utility. In mobile networks, game theory is applied in the form of matching games [128], where system players (e.g., users) have to be matched with network resources (e.g., base stations or resource blocks).

Three types of matching games can be used depending on the application scenario: 1) one-to-one matching, where each user can be matched with at most one resource (as in [129], which optimizes D2D communication in small cell scenarios); 2) many-to-one matching, where either multiple resources can be assigned to a single user (as in [130] for small cell resource allocation), or multiple users can be matched to a single resource (as in [131] for user-cell association); 3) many-to-many matching, where multiple users can be matched with multiple resource (as in [133] where videos are associated to caching servers).

E. Summary

This section (and Table VI) summarizes the main takeaways of this optimization handbook.

1) Convex Optimization Methods: These methods are often combined with time series analysis or ideal prediction. The main reason is that they are used to determine performance bounds when the solving time is not a system constraint. Thus, convex optimization is suggested as a benchmark for large scale prediction. This may have to be replaced by fast heuristics in case the optimization tool needs to work in real-time. An exception to this is LP for which very efficient algorithms exist that can compute a solution in polynomial time.

TABLE VI
ANTICIPATORY NETWORKING APPLICABILITY TO DIFFERENT NETWORK TYPES

Type	Features	Advantages	Challenges
5G Cellular	mm-waves Massive MIMO Cloud-RAN	Localization and tracking prediction Load space-time distribution Resource management	Channel models Amount of data
MANET	Variable topology Multi-hop communication Self-management	Routing improvement Load balancing	Infrastructure absence Distributed optimization Variable topology
Cognitive	Primary/Secondary users Sensing capabilities	Spectrum availability prediction Load prediction and management Transmission/Sensing ratio	Impact on models
D2D	Complex topology Multi-RAN	Interference management Resource allocation	Models complexity Interference
IoT	Mostly deterministic traffic High overhead Sparse communication Low-latency control loops	Prediction for compression Models for anomaly detection Overhead decrease	Amount of data and devices Scalability Constrained devices

In contrast, convex optimization methods should be preferred when dealing with high precision and continuous output. They require the complete dataset and show a reliability comparable to that of the used predictor.

2) Model Predictive Control: MPC combines prediction and optimization to minimize the control error by tuning both the prediction and the control parameters. Therefore, it can be coupled with any predictor. The main drawback of this approach is that, by definition, prediction and optimization cannot be decoupled and must be evaluated at each iteration. This makes the solution computationally very heavy and it is generally difficult to obtain real-time algorithms based on MPC. The close coupling between prediction and optimization makes it possible to adopt the method for any application for which a predictor can be designed with the only additional constraint being the execution time. Objectives and constraints are usually those imposed by the used predictor.

3) Markov Decision Processes: MDPs are characterized by a statistical description of the system state and they usually model the system evolution through probabilistic predictors. As such, they best fit to scenarios that show similar objective functions and constraints as those of probabilistic predictors. Thus, MDPs are the ideal choice when the optimization objective aims at obtaining stationary policies (i.e., policies that can be applied independently of the system time). This translates to low precision and high reliability. Moreover, even though they require a computationally heavy phase to optimize the policies, once the policies are obtained, fast algorithms can easily be applied.

4) Game Theory: Matching games prove to be effective solutions that, without struggling to compute an overly complex optimal configuration, let the system converge towards a stable equilibrium which satisfies all the players (i.e., no action can be taken to improve the utility of any player). These are the preferable solutions for those applications where the computational capability is a stringent constraint and where fairness is important for the system quality.

VI. APPLICABILITY OF ANTICIPATORY NETWORKING TO OTHER WIRELESS NETWORKS

So far this survey mainly focused on current cellular networks. In this section we analyze how different types of mobile wireless networks can take advantage of anticipatory networking solutions. Although each type would deserve a dedicated survey, in what follows we provide brief summaries of the distinctive features, the application scenarios, the expected benefits and the challenges related to the implementation of anticipatory networking for each of them. Table VI summarizes the discussion of this section.

A. 5G Cellular Networks

LTE and LTE-advanced represent the fourth generation of mobile cellular networks and, as it emerged from the analyses of the previous sections, they can already benefit from predictive optimization. Since the fifth generation is expected to improve on its predecessors in every aspect [166], not only is anticipatory networking applicable, but also it will provide even greater benefits.

1) Characteristics: The next generation of mobile cellular networks will provide faster communications, improved users QoE, shorter communication delays, higher reliability and improved energy savings. Among the solutions envisioned to realize these improvements, cell densification, mm-wave bands, massive MIMO, unified multi-technology frame structure and architecture and network function virtualization are the ones that are going to have a substantial impact on existing and future use case scenarios. In fact, a denser infrastructure is going to decrease the average time mobile users spend in a specific cell; the directionality of communications in higher portion of the spectrum will increase the importance of localization and tracking functionalities; while the increase of communicating elements and the de-localization of radio access functionalities are going to impact on channel models and network resource management.

2) Advantages: The performance of 5G cellular networks will strongly depend on their knowledge of the exact user positions (e.g., localization for mm-wave, resource management for network function virtualization). As a consequence, predictive solutions that provide the system with accurate information about users' current and future positions, trajectories, traffic profiles and content request probabilities are likely to be the most desirable aspects of anticipatory solutions.

For what concerns 5G applications, we believe network caching and cloud Radio Access Network (RAN) will also greatly benefit from this. In fact, the former can exploit prediction to decide which content to store in which specific part of the network to serve a given user profile, while the latter can, for instance, forecast when to instantiate a number of virtual machines to face an increase of the network traffic.

3) Challenges: The upcoming 5G technologies will also bring new challenges to the basic mechanisms of anticipatory networking. In particular, we see mm-wave, massive MIMO and cell densification as disruptive technologies for the current methods used for predictive optimization. In this regard, mm-waves channel model is going to impact how to forecast future signal quality and achievable data rates while network densification and massive MIMO will challenge the scalability of prediction techniques due to the sheer size of the information needed to describe and exchange them.

B. Mobile Ad Hoc Networks

Mobile Ad-hoc Networks (MANET) consist of mobile wireless devices connected to one another without a fixed infrastructure [167]. As a consequence, they share some characteristics with cellular networks but have some unique features due to the variable topology. These networks are the most practical form of communication when an infrastructure is absent or it has been compromised by a disruptive event.

1) Characteristics: The dynamic nature of MANETs causes the path between any two nodes to vary over time and require adaptive routing mechanisms that allow, on one hand, to maintain the connectivity among all the network nodes and, on the other hand, to balance the load in the different areas of the network. In addition, adaptive discovery and management functionalities are needed to allow new devices and services to be added to an existing network and to report problems and missing links/nodes. When a MANET extends over an area larger than the communication range of the devices, transmissions must be relayed from one node to another in order to allow messages to reach their destinations.

2) Advantages: Knowing nodes' positions in advance and being able to track their trajectories enable advanced routing functionalities: in fact, additional paths can be created before a missing link interrupts a route without waiting for a new discovery procedure to be performed. Also, routing tables can be readily adapted when shorter routes appear. In a similar way, management procedure can be enhanced by knowing in advance the traffic being produced by a given node or area of the network or by forecasting which service is going to be needed in a given part of the network.

3) Challenges: The absence of a fixed infrastructure is the main source of challenges that are distinctive of MANETs. For instance, it is not possible to have known databases collecting users' and devices' information to build prediction models nor centralized optimization services can be provided or they may suffer from delays in delivering solutions and/or information to the whole network. Moreover, the topology variability makes map-based prediction techniques difficult or impossible to apply.

C. Cognitive Radio Networks

CR networks consist of devices that exploit channels that are unused at specific locations and times [10], but that are usually allocated to primary users (i.e., users that can legitimately communicate using a given channel). CR devices are usually referred to as secondary users as their operations must not interfere with those performed by the primary users.

1) Characteristics: The main distinctive feature of CR devices is that they need to scan for primary users' activity before attempting any communication in order not to disrupt legitimate transmissions. This scanning/sensing activity decreases the amount of time secondary users' can spend on actual communications and, thus, it reduces their throughput. On the other hand, a CR network is usually able to build accurate spectrum occupancy models fusing the information coming from different devices.

2) Advantages: Prediction capabilities are already envisioned for CR networks, in fact, it is easily understandable that being able to predict when primary users are going to occupy their channel will decrease the amount of sensing needed to decide when a secondary user is allowed to transmit. Not only can spectrum occupancy maps be used to predict the upcoming channel state, but also, content information and predictive models available to primary users can be exploited by secondary users to reduce their interference probability. Therefore, allowing secondary users to access primary user information is profitable for both: if CR are able to improve their throughput by more precisely picking spectrum holes, primary users will be more protected from secondary interference.

3) Challenges: Although anticipatory CR can be seen as symbiotic to primary users, their operations introduce a non trivial feedback in the resulting system. In fact, those models that are valid when primary users operate only may be no longer valid when secondary users contribute. However, given that those models are usually built using information about primary users only, it will be impossible with the current techniques to create or modify prediction and optimization solutions that take into consideration secondary users. As such, the whole anticipatory infrastructure needs to account for CR in order to allow prediction-based schemes to work for primary and secondary users.

D. Device-to-Device

D2D communication refers to the use of direct communication between mobile phones to support the operations of a cellular network [168]. In addition, since D2D must not interfere with the regular cellular network operations it can be seen as secondary users to the main communications. Therefore, they share characteristics that are specific to MANETs and CR networks.

1) Characteristics: D2D communications are characterized by a complex topology where the usual star network overlies a mesh network. Also, the devices may use different RANs in the mesh network: for instance they can exploit the same cellular technology (inband) or other wireless solutions such as direct-WiFi.

2) Advantages: Given the similarities to MANETs and CRs, D2D communications can take advantage from anticipatory networking mostly to mitigate interference related problems and to improve the resource and power allocation.

3) Challenges: While we do not expect D2D communications to pose distinctive challenges to the implementation of anticipatory networking that are not listed in the previous sections, that will make the adoption of current prediction models less straightforward. In fact, prediction-based optimization and other anticipatory schemes will be made more complex due to the possible coexistence of multiple technologies and the primary/secondary interference and interactions, which will require to also predict D2D channels, in addition to primary.

E. Internet of Things

Nowadays, thanks to the miniaturization and the progressive decrease of computational and communicating chipsets, more and more ordinary objects are being equipped with micro-CPUs and are connected to the Internet [169]–[171]: in such a way smart cities and smart industries, among a variety of other enhanced scenarios, can be realized. The typical device in the Internet-of-Things (IoT) is capable of performing one or a set of measurements and/or actuations on the real world. They are usually constrained in their capabilities: for instance, they can be battery powered or equipped with low data rate radios or their computational power may be limited.

1) Characteristics: Due to the wide definition of the entities that populate the IoT, many of its features have been already described in the preceding subsections. For instance, IoT communications often involve D2D aspects, they can be CR if they are able to sense spectrum and they can be considered part of a MANET if they are mobile. However, the most unique features that are only present in IoT devices are that they involve Machine-to-Machine (M2M) type communication and that devices are typically constrained. Moreover, although the number of smart things is expected to grow exponentially in the next decade, their traffic is not going to grow as fast as that, e.g., the one generated by mobile cellular networks. In fact, IoT traffic is expected to be mainly due to monitoring, control and detection activities, which are characterized by limited throughput and almost deterministic transmission frequency.

2) Advantages: Anticipatory networking and prediction-based optimization can be applied to many aspects of the IoT. For instance, devices that harvest their energy from renewable sources may predict the source availability and optimize their operations according to that. Furthermore, data prediction models can be used to compress the data produced by devices by sending only the difference from the forecast or the same models can be used to identify anomalies or prevent disruptive events before they can cause serious problems. Finally, due to the almost deterministic periodicity of data production, their communication can be easily modeled and accounted for to mitigate their impact on the overall system.

3) Challenges: Scalability is one of the main challenges in IoT. In fact, due to the variety of device types, the difference in their capabilities, requirements and applications, the amount of information needed to represent and model the IoT is huge and the obtained benefits must more than compensate for the cost related to its realization. Moreover, the IoT is impacted by most of the challenges and problems discussed above for the other network types.

VII. ON THE IMPACT OF ANTICIPATORY NETWORKING ON THE PROTOCOL STACK

In this section, we address another important aspect of anticipatory networking solutions: where to implement them in the ISO/OSI protocol stack [172] and which layers contribute to their realizations.

A. Physical

We do not expect anticipatory networking solutions to modify how the physical layer is designed and managed. In fact, in order to apply prediction-based schemes, some form of interaction is required between two or more entities of the system. As a consequence, the physical layer, which defines how information is transferred to bits and wave-form [172], might provide different profiles to allow for predictive techniques to be applied in the higher layers, but will not directly implement any of them.

B. Data Link

The data link layer is the first entry point for predictive solutions. In particular, this layer implements Medium Access Control (MAC) functionalities. Therefore, resource management [42] and admission control [75] procedures are likely to greatly benefit from anticipatory optimization. Also, we envision that anticipatory networking to be even more important in next generation networks: in particular, channel estimation and beam steering solutions are going to be key for the success of mm-wave a massive MIMO communications [166].

C. Network

The network layer contains two of the functionalities that can benefit the most from prediction: routing and caching [54], [122]. In fact, by knowing users' mobility and traffic in advance it is possible to optimize routes and caching location to maximize network performance and save resources. For instance, it is possible to build alternative paths before the existing ones deteriorate and break and popular contents may be moved across the network according to where they will be requested with higher probability.

D. Transport

This layer is mainly concerned with end-to-end message delivery and the two most popular protocols are TCP and User Datagram Protocol. (UDP): the former guarantees reliable communications, while the latter is a lightweight best-effort solution. Anticipatory networking solutions are easily implemented here [31], [135], in particular, when error correction and retransmissions are driven by network metrics such as, among others, Round Trip Time (RTT) and Bit Error Rate (BER). Prediction models can be used to react to changes in the network conditions before they reach a disruptive state and recovery actions have to be taken. In addition, modern transport solutions, such as multipath-TCP, can exploit predictive optimization to manage the traffic flows along the different routes and improve the QoS.

E. Session, Presentation and Application

Since these layers are concerned with connection management between end-points (session), syntax mapping between different protocols (presentation) and interaction with users and software (application), they are the least preferable to implement anticipatory networking solutions. However, in order to allow applications to exploit predictive mechanisms, these three layers will act as a connection point to provide application with the needed context information and to allow them to configure the needed services and parameters for the application requirements. For instance, in Section III-A6 we described geographically-assisted video optimization [62], [77] where mobile phone applications modulated the request video bit rate to optimize the playback of the video itself, or geo-assisted applications [134] that exploits social and contextual information to enhance their services.

VIII. Issues, Challenges, and Research Directions

We conclude the paper by providing some insights on how anticipatory optimization will enable new 5G use cases and by detailing the open challenges of anticipatory networking in order to be successfully applied in 5G.

A. Context Related Analyses

1) Geographic Context: Geographic context is essential to achieve seamless service. Depending on the optimization objective, a mobility state can be defined with different granularity in multiple dimensions (location, time, speed, etc.). For example, for handover optimization it is sufficient to predict the staying time in the current serving cell and the next serving cell of the user. Medium to large spatial granularity such as cell ID or cell coverage area can be considered as a state, and a trajectory can be characterized by a discrete sequence of cell IDs over time. State-space models such as Markov chains, HMM and Kalman filters fit the system modeling, while requiring large training samples and considerable insight to make the model compact and tractable. An alternative is the variable-order Markov models, including a variety of lossless compression algorithms (some of the most used belong to Lempel-Ziv family), where Shannon's entropy measure is identified as a basis for comparing user mobility models. Such an information-theoretic approach enables adaptive online learning of the model, to reduce update paging cost. Moving from discrete to continuous models, which are applied to assist the prediction of other system metrics with high granularity, e.g., link gain or capacity, regression techniques are widely used. To enhance the prediction accuracy, a priori knowledge can be exploited to provide additional constraints on the content and form of the model, based on street layouts, traffic density, user profiles, etc. However, finding the right trade-off between the model accuracy and complexity is challenging. An effective solution is to decompose the state space and to introduce localized models, e.g., to use distinct models for weekdays and weekends, or urban and rural areas.

Although mobility prediction has been shown to be viable, it has not been widely adopted in practical systems. This is because, unlike location-aware applications with users' permission to use their location information, mobile service providers must not violate the privacy and security of mobile users. To facilitate the next generation of user-centric networks, new interaction protocols and platforms need to be developed for enabling more user-friendly agreements on the data usage between the service providers and the mobile users.

Furthermore, next generation wireless networks introduce ultra-dense small cells and high frequencies such as mmWaves. The transmission range gets shorter and transmission often occurs in line-of-sight conditions. Thus, 2D geographic context with a coarse level of accuracy is not sufficient to fully utilize the future radio techniques and resources. This trend opens the door for new research directions in inference and prediction of 3D geographic context, by utilizing advanced feedback from sensors in user equipments such as accelerometers, magnetometers, and gyroscopes.

2) Link Context: When predicting link context, i.e., channel quality and its parameters, linear time series models have the potential to provide the best tradeoff between performance and complexity. When the channel changes slowly, e.g., because users are static or pedestrian, it is convenient to exploit the temporal correlation of historic measurements of the users' channel and implement linear auto-regressive prediction. This can be quite accurate for very short prediction horizons and at the same time simple enough to be implemented in real time systems. Kalman filters can also be used to track errors and their variance, based on previous measurements, thus handling uncertainties. However, time series and linear models are not robust to fast changes. Therefore, in high mobility scenarios, more complex models are needed. One possible approach is to exploit the spatio-temporal correlation between location and channel quality. By combining the prediction of the channel qualities with the prediction of the user's trajectory, regression analysis, e.g., SVMs, can be employed to build accurate radio maps to estimate the long term average channel quality, which accounts for pathloss and slow fading, but neglects fast fading variations. Ideally, one should have two predictions available: a very accurate short term prediction and an approximate long term prediction.

Usually, such prediction is exploited to optimize the scheduling, i.e., resource allocation over time or frequency. Convex and linear optimization are often used when prediction is assumed to be perfect. In contrast, Markov models are applied when a probabilistic forecasting is available. Despite the great benefits that link context can potentially bring to resource (and more generally network) optimization, today's networks do not yet have the proper infrastructure to collect, share, process and distribute link context. Furthermore, proper methods are needed not only to gather data from users, but also, to discard irrelevant or redundant measurements as well as to handle sparsity or gaps in the collected data.

3) Traffic Context: Traffic and throughput prediction has a concrete impact on the optimization of different services of different networks at different time scales.

Network-wide and for long time scales, linear time series models are already used to predict the macroscopic traffic patterns of mobile radio cells for medium/long-term management and optimization of the radio resources. At faster time scales and for specific radio cells or groups of radio cells, the probabilistic forecasting of the upcoming traffic, e.g., by using Markovian models, can be exploited to solve short-term problems including the radio resource allocation among users and the cell assignment problem.

Throughput prediction tools are then naturally coupled with video streaming services in mobile radio networks which have embedded rate adaptation capabilities. In this context, a good practice is to use simple yet effective look-ahead video throughput predictors based on time windows which are often coupled with clustering approaches to group similar video sessions. Deep learning techniques are also proposed to predict the throughput of video sessions, which offer improved performance at the price of a much higher complexity.

The data coming from traffic/throughput prediction can be effectively coupled with application/scenario-specific optimization frameworks. When targeting network-wide efficiency, centralized optimization approaches seem to be superior and more widely used. As an example, the problem of radio resource allocation in mobile radio networks is effectively representable and solvable though convex optimization techniques in semi-real-time scenario. In contrast, when the optimization has to be performed with the granularity of the technology-specific time slot, sub-optimal heuristics are preferable. Besides resorting to optimization approaches, control theoretic modeling is extremely powerful in all those cases where the optimization objective includes traffic (and queue) stability.

4) Social Context: We can conclude that leveraging the social context of data transmission results in gains for proactive caching of multimedia content and can improve resource allocation by predicting the social behavior of users. For the former, determining the popularity of content plays a crucial role. Collaborative filtering is a well-known approach for this purpose. However, due to the heavy tail nature of content popularity, trying to use this kind of models for a broad class of content will usually not lead to good results. However, for more specific and limited classes of content, i.e., localized advertisement, where a particular item is likely to be requested by a large number of users, popularity prediction is an appealing solution. In general, proactive caching requires that content is stored on caches close to the edge network in order not to put excessive load on the core network. For optimizing resource allocation using social behavior, the social interaction of different users can be used to create social graphs that determine the level of activity of each user and thereby make it possible to predict the amount of resources each user will need. Network utility maximization and heuristic methods are the most popular techniques for this context. Due to the complexity of modeling the social behavior of users, they are useful for wireless networks that either expose a great deal of measurable social interaction (device-to-device communication, dense cellular networks with small cells, local wireless networks in a sports stadium), or when resources are very scarce.

B. Anticipation-Enabled Use Cases

Future networks are envisioned to cater to a large variety of new services and applications. Broadband access in dense areas, massive sensor networks, tactile Internet and ultra-reliable communications are only a few of the use cases detailed in [173]. The network capabilities of today's systems (i.e., 4G systems) are not able to support such requirements. Therefore, 5G systems will be designed to guarantee an efficient and flexible use (and sharing) of wireless resources, supported by a native software defined network and/or network function virtualization architecture [173]. Big data analysis and context awareness are not only enablers for new value added services but, combined with the power of anticipatory optimization, can play a role in the 5G technology.

1) Mobility Management: Network densification will be used in 5G systems in order to cope with the tremendous growth of traffic volume. As a drawback, mobility management will become more difficult. Additionally, it is foreseen that mobility in 5G will be on-demand [173], i.e., provided for and customized to the specific service that needs it. In this sense, being able to predict the user's context (e.g., requested service) and his mobility behavior can be extremely useful in order to speed up handover procedures and to enable seamless connectivity. Furthermore, since individual mobility is highly social, social context and mobility information will be jointly used to perform predictions for a group of socially related individuals.

2) Network Sharing: 5G systems will support resource and network sharing among different stakeholders, e.g., operators, infrastructure providers, service providers. The effectiveness of such sharing mechanisms relies on the ability of each player to predict the evolution of his own network, e.g., expected network load, anticipated user's link quality and prediction of the requested services. Wireless sharing mechanisms can strongly benefit from the added value provided by anticipation, especially when prediction is available at fine granularity, e.g., in a multi-operator scheduler [174].

3) Extreme Real-Time Communications: Tactile Internet is only one of the applications that will require a very low latency (i.e., in the order of some milliseconds). Allocating resources and guaranteeing such low end-to-end delay will be very challenging. 5G systems will support such requirements by means of a new physical layer (e.g., a new air interface). However, this will not be enough if not combined with context information used to prioritize control information (e.g., used to move virtual or real objects in real time) over content [175]. Knowledge about the information that is transmitted and its specific requirements will be crucial in order to assign priorities and meet the expected quality-of-experience in a combined effort of physical and higher layers.

4) Ultra-Reliable Communications: Reliability is mentioned in several 5G white papers, e.g., in [173], as necessary prerequisite for lifeline communications and e-health services, e.g., remote surgery. A recent work [176] proposed a quantified definition of reliability in wireless access networks. As outlined here, a posteriori evaluation of the achieved reliability is not enough in order to meet the expected target, which in some cases is as high as 99.999%. To this end, it is mandatory to design resource allocation mechanisms that account for (and are able to anticipate the impact on) reliability in advance.

C. Open Challenges

While the literature surveyed so far clearly points out how anticipatory networking can enhance current networks, this section discusses several problems that need to be solved for its wider adoption. In particular, we identified four functionalities that are going to play an important role in the adoption of anticipatory networking in 5G networks:

- **Measurements and information collection:** in order to provide means to obtain and share context information, future networks need to provide trusted mechanisms to manage the information exchange.
- **Data analysis and prediction:** information databases need interoperable procedures to make sure that processing and forecasting tools are usable with many possible information sources .
- **Optimization and decision making:** data and procedures are then exploited to derive system management policies.
- **Execution:** finally, in contrast to current procedures, anticipatory execution engines need to take into account the impact of the decisions made in the past and re-evaluate their costs and rewards in hindsight of the actual evolution of the system.

For instance, scheduling and load balancing are two processes that greatly profit from anticipatory networking and cannot be realized without a comprehensive integration of the four aforementioned functionalities in future generation networks. The realization of these functionalities poses the following important challenges.

1) Privacy and Security: In our opinion, one of the main hindrances for anticipatory networking to become part of next generation networks is related to how users feel about sharing data and being profiled. While voluntarily sharing personal information has become a daily habit, many disapprove that companies create profiles using their data [177]. In a similar way, there might be a strong resistance against a new technology that, even though in an anonymous way, collects and analyzes users' behavior to anticipate users' decisions. Standards and procedures need to be studied to enforce users' privacy, data anonymity and an adequate security level for information storage. In addition, data ownership and control need to be defined and regulated in order to allow users and providers to interact in a trusted environment, where the former can decide the level of information disclosure and the latter can operate within shared agreements.

2) Network Functions and Interfaces: Many of the applications that are likely to benefit from anticipatory networking capabilities (i.e., decision making and execution) require unprecedented interactions among information producers, analyzers and consumers. A simple example is provided by predictive media streaming optimizers, which need to obtain content information from the related database and user streaming information from the user and/or the network operator. This information is then analyzed and fed to a streaming provider that optimizes its service accordingly. While ad hoc services can be realized exploiting the current networking functionalities, next generation applications, such as the extreme real-time communications mentioned above, will greatly benefit from a tighter coupling between context information and communication interfaces. We believe that the potential of anticipatory functionalities can be used in communication system and they could be applied to other domains, such as public transportation and smart city management.

3) Next Generation Architecture: 5G networks are currently being discussed and, while much attention is paid to increasing the network capacity and virtualizing the network functions, we believe that the current infrastructure should be enhanced with repositories for context information and application profiles [178] to assist the realization of novel predictive applications. As per the previous concerns above, sharing sensible information, even in an anonymized way, will require particular care in terms of users' privacy and database accessibility. We believe that anticipatory networking can potentially improve every kind of mobile networks: cellular networks will likely be the first to exploit this paradigm, because they already own the information needed to enable the predictive frameworks and it is only a matter of time and regulations to make it a reality. Once it will be integrated in cellular networks, other systems, such as public WiFi deployments, device-to-device solutions and the Internet of Things, will be able to participate in the infrastructure to exploit forecasting functionalities; in particular, we believe this will be applied to smart cities and multi-modal transportation.

4) Impact of Prediction Errors: When making and using predictions, one should carefully estimate its accuracy, which is itself a challenge. It might be potentially more harmful to use a wrong prediction than not using prediction at all. Usually, a good accuracy can be obtained for a short prediction horizon, which, however, should not be too short, otherwise the optimization algorithms cannot benefit from it. Therefore, a good balance between prediction horizon and accuracy must be found in order to provide gains. In contrast, over medium/long term periods, metrics can usually be predicted in terms of statistical behavior only. Furthermore, to build robust algorithms that are able to deal with uncertainties, proper prediction error models should be derived. In the existing literature, uncertainties are mainly modeled as Gaussian random variables. Despite the practicability of such an assumption, more complex error models should be derived to take into account the source (e.g., location and/or channel quality) as well as the cause (e.g., GPS accuracy and/or fast fading effect) of errors.

IX. Conclusion

This survey analyzed the literature on anticipatory networking for mobile networks. We provided a thorough analysis of application scenarios categorized by the contextual information used to build the predictive framework. The most relevant prediction and optimization techniques adopted in the literature have been described and commented in two handbooks that have the twofold objective of supporting researchers to advance in the field and providing standardization and regulation bodies with a common ground on anticipatory networking solutions. While the core of this survey is devoted to mobile cellular networks, we also analyzed applicability and advantages of anticipatory networking solution to other types of wireless networks and at the different layers of the protocol stack. Finally, we analyzed benefits and disadvantages of the proposed solutions, the most promising application scenarios for 5G networks, and the challenges that are yet to be faced to adopt anticipatory networking paradigms.

To conclude, while the literature reviewed in this works suggests that anticipatory networking is a quite mature approach to improve the performance of mobile networks, we believe that issues (mainly at the system level) still need to be solved to realize its potential. In particular, most of the work which has been evaluated in this survey tends to focus on the benefit of anticipation, while overlooking possible problems and disadvantages in the anticipatory networking framework.

All the main components of anticipatory networking, the context database and the prediction/anticipation intelligence, must be effectively integrated into the mobile network architecture which poses challenges at different levels. First, new interfaces and communication paradigms must be defined for data collection from both end users and sources external to the mobile network itself; second, the management of the context databases brings an additional burden in terms of required bandwidth and processing power for several network elements which may lead to scalability issues as well as security and privacy concerns. To this extent, a thorough and comprehensive cost-benefit analysis for specific anticipatory networking scenarios is, in our opinion, a required next step for the research in the field.

X. List of Acronyms

ANN	Artificial Neural Network
AR	AutoRegressive
ARIMA	AutoRegressive Integrated and Moving Average
ARMA	AutoRegressive and Moving Average
ATM	Asynchronous Transfer Mode
BER	Bit Error Rate
CCN	Content Centric Network
CF	Collaborative Filtering
ConvOpt	Convex Optimization
CR	Cognitive Radio
CSI	Channel State Information
CTM	Continuous Time Markov
CTMC	Continuous Time Markov Chain
D2D	device-to-device
DASH	Dynamic Adaptive Streaming over HTTP
DTMC	Discrete Time Markov Chain
ELM	Extreme Learning Machine
FTP	File Transfer Protocol
GARCH	Generalized AutoRegressive Conditionally Heteroskedastic
GP	Gaussian Process
GPS	Global Positioning System
HMM	Hidden Markov Models
HTTP	Hypertext Transfer Protocol
ID	identity
ILP	Integer Linear Programming
IoT	Internet-of-Things
KKF	Kriged Kalman Filter
LTE	Long Term Evolution
LP	Linear Programming
LZ	Lempel-Ziv
M2M	Machine-to-Machine
MA	Moving Average
MAC	Medium Access Control
MANET	Mobile Ad-hoc Networks
MC	Markov Chain
MILP	Mixed-Integer Linear Programming
MNLP	Mixed Non-Linear Program
MPC	Model Predictive Control
MDP	Markov Decision Process
PF	Proportionally Fair
QoE	Quality-of-Experience
QoS	Quality-of-Service
RAN	Radio Access Network
REM	Radio Environment Map
RTT	Round Trip Time
SVM	Support Vector Machine
TCP	Transmission Control Protocol
TCP	Transport Control Protocol
UDP	User Datagram Protocol.

References

[1] K. Zheng *et al.*, "Big data-driven optimization for mobile networks toward 5G," *IEEE Netw.*, vol. 30, no. 1, pp. 44–51, Jan./Feb. 2016.

[2] P. Makris, D. N. Skoutas, and C. Skianis, "A survey on context-aware mobile and wireless networking: On networking and computing environments' integration," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 362–386, 1st Quart., 2013.

[3] V. Pejovic and M. Musolesi, "Anticipatory mobile computing: A survey of the state of the art and research challenges," *ACM Comput. Surveys*, vol. 47, no. 3, 2015, Art. no. 47.

[4] S. Boucheron, O. Bousquet, and G. Lugosi, "Theory of classification: A survey of some recent advances," *ESAIM Probab. Stat.*, vol. 9, pp. 323–375, Nov. 2005.

[5] Y. Liu and J. Y. B. Lee, "An empirical study of throughput prediction in mobile data networks," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, San Diego, CA, USA, 2015, pp. 1–6.

[6] T. T. T. Nguyen and G. Armitage, "A survey of techniques for Internet traffic classification using machine learning," *IEEE Commun. Surveys Tuts.*, vol. 10, no. 4, pp. 56–76, 4th Quart., 2008.

[7] L. Jin, Y. Chen, T. Wang, P. Hui, and A. V. Vasilakos, "Understanding user behavior in online social networks: A survey," *IEEE Commun. Mag.*, vol. 51, no. 9, pp. 144–150, Sep. 2013.

[8] S. Baraković and L. Skorin-Kapov, "Survey and challenges of QoE management issues in wireless networks," *Hindawi J. Comput. Netw. Commun.*, vol. 2013, Mar. 2013, Art. no. 165146. AQ6

[9] M. Höyhtyä *et al.*, "Spectrum occupancy measurements: A survey and use of interference maps," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 4, pp. 2386–2414, 4th Quart., 2016.

[10] Y. Chen and H.-S. Oh, "A survey of measurement-based spectrum occupancy modeling for cognitive radios," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 848–859, 1st Quart., 2016.

[11] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.
[12] X. Lu, E. Wetter, N. Bharti, A. J. Tatem, and L. Bengtsson, "Approaching the limit of predictability in human mobility," *Nature Sci. Rep.*, vol. 3, Oct. 2013, Art. no. 2923.
[13] Y. Jiang, D. C. Dhanapala, and A. P. Jayasumana, "Tracking and prediction of mobility without physical distance measurements in sensor networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Budapest, Hungary, 2013, pp. 1845–1850.
[14] L. Ghouti, T. R. Sheltami, and K. S. Alutaibi, "Mobility prediction in mobile ad hoc networks using extreme learning machines," *Proc. Comput. Sci.*, vol. 19, pp. 305–312, Dec. 2013.
[15] X. Chen, F. Mériaux, and S. Valentin, "Predicting a user's next cell with supervised learning based on channel states," in *Proc. IEEE Signal Process. Adv. Wireless Commun. (SPAWC)*, Darmstadt, Germany, 2013, pp. 36–40.
[16] H. Xiong *et al.*, "MPaaS: Mobility prediction as a service in telecom cloud," *Springer Inf. Syst. Front.*, vol. 16, no. 1, pp. 59–75, 2014.
[17] J.-K. Lee and J. C. Hou, "Modeling steady-state and transient behaviors of user mobility: Formulation, analysis, and application," in *Proc. ACM Int. Symp. Mobile Ad Hoc Netw. Comput. (MobiHoc)*, Florence, Italy, 2006, pp. 85–96.
[18] H. Abu-Ghazaleh and A. S. Alfa, "Application of mobility prediction in wireless networks using Markov renewal theory," *IEEE Trans. Veh. Technol.*, vol. 59, no. 2, pp. 788–802, Feb. 2010.
[19] D. Barth, S. Bellahsene, and L. Kloul, "Mobility prediction using mobile user profiles," in *Proc. IEEE Model. Anal. Simulat. Comput. Telecommun. Syst. (MASCOTS)*, Singapore, 2011, pp. 286–294.
[20] D. Barth, S. Bellahsene, and L. Kloul, "Combining local and global profiles for mobility prediction in LTE femtocells," in *Proc. ACM Model. Anal. Simulat. Wireless Mobile Syst. (MSWiM)*, Paphos, Cyprus, 2012, pp. 333–342.
[21] G. Gidófalvi and F. Dong, "When and where next: Individual mobility prediction," in *Proc. ACM SIGSPATIAL Int. Workshop Mobile Geographic Inf. Syst.*, Redondo Beach, CA, USA, 2012, pp. 57–64.
[22] Y. Chon, N. D. Lane, Y. Kim, F. Zhao, and H. Cha, "Understanding the coverage and scalability of place-centric crowdsensing," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. (Ubicomp)*, Zürich, Switzerland, 2013, pp. 3–12.
[23] Y. Chon, H. Shin, E. Talipov, and H. Cha, "Evaluating mobility models for temporal prediction with high-granularity mobility data," in *Proc. IEEE Pervasive Comput. Commun. (PerCom)*, Lugano, Switzerland, 2012, pp. 206–212.
[24] Y. Chon, E. Talipov, H. Shin, and H. Cha, "SmartDC: Mobility prediction-based adaptive duty cycling for everyday location monitoring," *IEEE Trans. Mobile Comput.*, vol. 13, no. 3, pp. 512–525, Mar. 2014.
[25] Y. Chon, Y. Kim, H. Shin, and H. Cha, "Adaptive duty cycling for place-centric mobility monitoring using zero-cost information in smartphone," *IEEE Trans. Mobile Comput.*, vol. 13, no. 8, pp. 1694–1706, Aug. 2014.
[26] Y. Chon, E. Talipov, H. Shin, and H. Cha, "Mobility prediction-based smartphone energy optimization for everyday location monitoring," in *Proc. ACM Conf. Embedded Netw. Sensor Syst. (SenSys)*, Seattle, WA, USA, 2011, pp. 82–95.
[27] I. F. Akyildiz and W. Wang, "The predictive user mobility profile framework for wireless multimedia networks," *IEEE/ACM Trans. Netw.*, vol. 12, no. 6, pp. 1021–1035, Dec. 2004.
[28] S. Scellato, M. Musolesi, C. Mascolo, V. Latora, and A. T. Campbell, "Nextplace: A spatio-temporal prediction framework for pervasive systems," in *Pervasive Computing*, vol. 6696. Heidelberg, Germany: Springer, 2011, pp. 152–169. AQ7
[29] M. De Domenico, A. Lima, and M. Musolesi, "Interdependence and predictability of human mobility and social interactions," *Elsevier Pervasive Mobile Comput.*, vol. 9, no. 6, pp. 798–807, 2013.
[30] P. Fazio, M. Tropea, F. De Rango, and M. Voznak, "Pattern prediction and passive bandwidth management for hand-over optimization in QoS cellular networks with vehicular mobility," *IEEE Trans. Mobile Comput.*, vol. 15, no. 11, pp. 2809–2824, Nov. 2016.
[31] H. Abou-Zeid, H. S. Hassanein, Z. Tanveer, and N. AbuAli, "Evaluating mobile signal and location predictability along public transportation routes," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, New Orleans, LA, USA, 2015, pp. 1195–1200.
[32] J. Yang and Z. Fei, "Broadcasting with prediction and selective forwarding in vehicular networks," *Hindawi Int. J. Distrib. Sensor Netw.*, vol. 2013, Dec. 2013, Art. no. 309041.
[33] A. Sridharan and J. Bolot, "Location patterns of mobile users: A large-scale study," in *Proc. IEEE INFOCOM*, Turin, Italy, 2013, pp. 1007–1015.
[34] J. Froehlich and J. Krumm, "Route prediction from trip observations," SAE, Troy, MI, USA, Tech. Rep., 2008. AQ8
[35] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti, "WhereNext: A location predictor on trajectory pattern mining," in *Proc. ACM Int. Conf. Knowl. Disc. Data Min. (SIGKDD)*, Paris, France, 2009, pp. 637–646.
[36] *GeoPKDD: Geographic Privacy-Aware Knowledge Discovery and Delivery 2005–2008*. [Online]. Available: http://www.geopkdd.eu AQ9
[37] N. Bui, F. Michelinakis, and J. Widmer, "A model for throughput prediction for mobile users," in *Proc. Eur. Wireless*, Barcelona, Spain, 2014, pp. 1–6.
[38] Q. Liao, S. Valentin, and S. Stańczak, "Channel gain prediction in wireless networks based on spatial-temporal correlation," in *Proc. IEEE Signal Process. Adv. Wireless Commun. (SPAWC)*, Stockholm, Sweden, 2015, pp. 400–404.
[39] (2004). *MOMENTUM, MOdels and siMulations for nEtwork plaNning and conTrol of UMts*. [Online]. Available: http://www.zib.de/momentum
[40] W. Wanalertlak *et al.*, "Behavior-based mobility prediction for seamless handoffs in mobile wireless networks," *Springer Wireless Netw.*, vol. 17, no. 3, pp. 645–658, 2011.
[41] H. Riiser, T. Endestad, P. Vigmostad, C. Griwodz, and P. Halvorsen, "Video streaming using a location-based bandwidth-lookup service for bitrate planning," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 8, no. 3, pp. 1–19, 2012.
[42] Z. Lu and G. De Veciana, "Optimizing stored video delivery for mobile networks: The value of knowing the future," in *Proc. IEEE INFOCOM*, Turin, Italy, 2013, pp. 2706–2714.
[43] H. Abou-Zeid, H. S. Hassanein, and S. Valentin, "Optimal predictive resource allocation: Exploiting mobility patterns and radio maps," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, Atlanta, GA, USA, 2013, pp. 4877–4882.
[44] R. Margolies *et al.*, "Exploiting mobility in proportional fair cellular scheduling: Measurements and algorithms," in *Proc. IEEE INFOCOM*, Toronto, ON, Canada, 2014, pp. 1339–1347.
[45] V. A. Siris and D. Kalyvas, "Enhancing mobile data offloading with mobility prediction and prefetching," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 17, no. 1, pp. 22–29, 2013.
[46] J. Hao, R. Zimmermann, and H. Ma, "Gtube: Geo-predictive video streaming over http in mobile environments," in *Proc. ACM Multimedia Syst. Conf. (MMSys)*, Singapore, 2014, pp. 259–270.
[47] X. Tie, A. Seetharam, A. Venkataramani, D. Ganesan, and D. L. Goeckel, "Anticipatory wireless bitrate control for blocks," in *Proc. ACM Conf. Emerg. Netw. Exp. Technol. (CoNEXT)*, Tokyo, Japan, 2011, Art. no. 9.
[48] M. Piacentini and F. Rinaldi, "Path loss prediction in urban environment using learning machines and dimensionality reduction techniques," *Springer Comput. Manag. Sci.*, vol. 8, no. 4, pp. 371–385, 2011.
[49] E. Dall'Anese, S.-J. Kim, and G. B. Giannakis, "Channel gain map tracking via distributed Kriging," *IEEE Trans. Veh. Technol.*, vol. 60, no. 3, pp. 1205–1211, Mar. 2011.
[50] S. Yin, D. Chen, Q. Zhang, and S. Li, "Prediction-based throughput optimization for dynamic spectrum access," *IEEE Trans. Veh. Technol.*, vol. 60, no. 3, pp. 1284–1289, Mar. 2011.
[51] S. J. Tarsa, M. Comiter, M. B. Crouse, B. McDanel, and H. T. Kung, "Taming wireless fluctuations by predictive queuing using a sparse-coding link-state model," in *Proc. ACM Int. Symp. Mobile Ad Hoc Netw. Comput. (MobiHoc)*, Hangzhou, China, 2015, pp. 287–296.
[52] M. Kasparick, R. L. G. Cavalcante, S. Valentin, S. Stanczak, and M. Yukawa, "Kernel-based adaptive online reconstruction of coverage maps with side information," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5461–5473, Jul. 2015.
[53] A. J. Nicholson and B. D. Noble, "Breadcrumbs: Forecasting mobile connectivity," in *Proc. ACM Int. Conf. Mobile Comput. Netw. (MobiCom)*, San Francisco, CA, USA, 2008, pp. 46–57.
[54] S. Naimi, A. Busson, V. Vèque, L. B. H. Slama, and R. Bouallegue, "Anticipation of ETX metric to manage mobility in ad hoc wireless networks," in *Ad-Hoc, Mobile, and Wireless Networks*. Cham, Switzerland: Springer, 2014, pp. 29–42.
[55] L. S. Muppirisetty, T. Svensson, and H. Wymeersch, "Spatial wireless channel prediction under location uncertainty," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1031–1044, Feb. 2016.

[56] M. Fröhle, L. S. Muppirisetty, and H. Wymeersch, "Channel gain prediction for multi-agent networks in the presence of location uncertainty," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Shanghai, China, 2016, pp. 3911–3915.

[57] L. S. Muppirisetty, J. Tadrous, A. Eryilmaz, and H. Wymeersch, "On proactive caching with demand and channel uncertainties," in *Proc. IEEE Conf. Commun. Control Comput. (Allerton)*, Monticello, IL, USA, 2015, pp. 1174–1181.

[58] N. Bui and J. Widmer, "Mobile network resource optimization under imperfect prediction," in *Proc. IEEE World Wireless Mobile Multimedia Netw. (WoWMoM)*, Boston, MA, USA, 2015, pp. 1–9.

[59] X. Wang, M. Chen, T. T. Kwon, L. T. Yang, and V. C. M. Leung, "AMES-Cloud: A framework of adaptive mobile video streaming and efficient social video sharing in the clouds," *IEEE Trans. Multimedia*, vol. 15, no. 4, pp. 811–820, Jun. 2013.

[60] W. Bao and S. Valentin, "Bitrate adaptation for mobile video streaming based on buffer and channel state," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., 2015, pp. 3076–3081.

[61] A. Seetharam *et al.*, "On managing quality of experience of multiple video streams in wireless networks," *IEEE Trans. Mobile Comput.*, vol. 14, no. 3, pp. 619–631, Mar. 2015.

[62] S. A. Hosseini, F. Fund, and S. S. Panwar, "(Not) yet another policy for scalable video delivery to mobile users," in *Proc. ACM Int. Workshop Mobile Video (MoVid)*, Portland, OR, USA, 2015, pp. 17–22.

[63] E. Kurdoglu *et al.*, "Real-time bandwidth prediction and rate adaptation for video calls over cellular networks," in *Proc. ACM Int. Conf. Multimedia Syst. (MMSys)*, Klagenfurt, Austria, 2016, Art. no. 12.

[64] Z. Liu and Y. Wei, "Hop-by-hop adaptive video streaming in content centric network," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, 2016, pp. 1–7.

[65] M. Dräxler, J. Blobel, and H. Karl, "Anticipatory download scheduling in wireless video streaming with uncertain data rate prediction," in *Proc. IFIP Wireless Mobile Netw. Conf. (WMNC)*, Munich, Germany, 2015, pp. 136–143.

[66] D. Tsilimantos, A. Nogales-Gómez, and S. Valentin, "Anticipatory radio resource management for mobile video streaming with linear programming," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, 2016, pp. 1–6.

[67] R. Atawia, H. Abou-Zeid, H. S. Hassanein, and A. Noureldin, "Robust resource allocation for predictive video streaming under channel uncertainty," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, Austin, TX, USA, 2014, pp. 4683–4688.

[68] T. Mangla, N. Theera-Ampornpunt, M. Ammar, E. Zegura, and S. Bagchi, "Video through a crystal ball: Effect of bandwidth prediction quality on adaptive streaming in mobile environments," in *Proc. ACM Int. Workshop Mobile Video (MoVid)*, Klagenfurt, Austria, 2016, Art. no. 1.

[69] R. Atawia, H. Abou-Zeid, H. S. Hassanein, and A. Noureldin, "Chance-constrained QoS satisfaction for predictive video streaming," in *Proc. IEEE Local Comput. Netw. (LCN)*, 2015, pp. 253–260.

[70] R. Atawia, H. Abou-Zeid, H. S. Hassanein, and A. Noureldin, "Joint chance-constrained predictive resource allocation for energy-efficient video streaming," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1389–1404, May 2016.

[71] E. Hossain and V. K. Bhargava, "Link-level traffic scheduling for providing predictive QoS in wireless multimedia networks," *IEEE Trans. Multimedia*, vol. 6, no. 1, pp. 199–217, Feb. 2004.

[72] H. Abou-Zeid, H. S. Hassanein, and S. Valentin, "Energy-efficient adaptive video transmission: Exploiting rate predictions in wireless networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 2013–2026, Jun. 2014.

[73] H. Abou-Zeid and H. S. Hassanein, "Efficient lookahead resource allocation for stored video delivery in multi-cell networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Istanbul, Turkey, 2014, pp. 1909–1914.

[74] N. Bui, I. Malanchini, and J. Widmer, "Anticipatory admission control and resource allocation for media streaming in mobile networks," in *Proc. ACM Model. Anal. Simulat. Wireless Mobile Syst. (MSWiM)*, Cancún, Mexico, 2015, pp. 255–262.

[75] N. Bui, S. Valentin, and J. Widmer, "Anticipatory quality-resource allocation for multi-user mobile video streaming," in *Proc. IEEE Workshop Commun. Netw. Techn. Contemp. Video (CNCTV)*, Hong Kong, 2015, pp. 245–250.

[76] M. Dräxler and H. Karl, "Cross-layer scheduling for multi-quality video streaming in cellular wireless networks," in *Proc. IEEE Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Sardinia, Italy, 2013, pp. 1181–1186.

[77] M. Dräxler, J. Blobel, P. Dreimann, S. Valentin, and H. Karl, "SmarterPhones: Anticipatory download scheduling for wireless video streaming," in *Proc. IEEE Int. Conf. Workshops Netw. Syst. (NetSys)*, 2015, pp. 1–8.

[78] S. Valentin, "Anticipatory resource allocation for wireless video streaming," in *Proc. IEEE Int. Conf. Commun. Syst. (ICCS)*, 2014, pp. 107–111.

[79] X. K. Zou *et al.*, "Can accurate predictions improve video streaming in cellular networks?" in *Proc. ACM Int. Workshop Mobile Comput. Syst. Appl. (HotMobile)*, 2015, pp. 57–62.

[80] X. Xing, T. Jing, W. Cheng, Y. Huo, and X. Cheng, "Spectrum prediction in cognitive radio networks," *IEEE Wireless Commun.*, vol. 20, no. 2, pp. 90–96, Apr. 2013.

[81] Z. Wei, Q. Zhang, Z. Feng, W. Li, and T. A. Gulliver, "On the construction of radio environment maps for cognitive radio networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Shanghai, China, 2013, pp. 4504–4509.

[82] H. B. Yilmaz, T. Tugcu, F. Alagöz, and S. Bayhan, "Radio environment map as enabler for practical cognitive radio networks," *IEEE Commun. Mag.*, vol. 51, no. 12, pp. 162–169, Dec. 2013.

[83] K. M. Thilina, K. W. Choi, N. Saquib, and E. Hossain, "Machine learning techniques for cooperative spectrum sensing in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 11, pp. 2209–2221, Nov. 2013.

[84] Z. Khan, J. J. Lehtomäki, L. A. DaSilva, E. Hossain, and M. Latva-Aho, "Opportunistic channel selection by cognitive wireless nodes under imperfect observations and limited memory: A repeated game model," *IEEE Trans. Mobile Comput.*, vol. 15, no. 1, pp. 173–187, Jan. 2016.

[85] Y. Saleem and M. H. Rehmani, "Primary radio user activity models for cognitive radio networks: A survey," *J. Netw. Comput. Appl.*, vol. 43, pp. 1–16, Aug. 2014.

[86] M. Monemi, M. Rasti, and E. Hossain, "Characterizing feasible interference region for underlay cognitive radio networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2015, pp. 7603–7608.

[87] M. Monemi, M. Rasti, and E. Hossain, "On characterization of feasible interference regions in cognitive radio networks," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 511–524, Feb. 2016.

[88] M. Ozger and O. B. Akan, "On the utilization of spectrum opportunity in cognitive radio networks," *IEEE Commun. Lett.*, vol. 20, no. 1, pp. 157–160, Jan. 2016.

[89] F. Akhtar, M. H. Rehmani, and M. Reisslein, "White space: Definitional perspectives and their role in exploiting spectrum opportunities," *Telecommun. Policy*, vol. 40, no. 4, pp. 319–331, 2016.

[90] A. A. Khan, M. H. Rehmani, and M. Reisslein, "Cognitive radio for smart grids: Survey of architectures, spectrum sensing mechanisms, and networking protocols," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 860–898, 1st Quart., 2016.

[91] S. H. R. Bukhari, M. H. Rehmani, and S. Siraj, "A survey of channel bonding for wireless networks and guidelines of channel bonding for futuristic cognitive radio sensor networks," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 924–948, 2nd Quart., 2016.

[92] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das, "Understanding traffic dynamics in cellular data networks," in *Proc. IEEE INFOCOM*, 2011, pp. 882–890.

[93] M. Z. Shafiq, L. Ji, A. X. Liu, and J. Wang, "Characterizing and modeling Internet traffic dynamics of cellular devices," in *Proc. ACM Joint Int. Conf. Meas. Model. Comput. Syst. (SIGMETRICS)*, San Jose, CA, USA, 2011, pp. 305–316.

[94] Z. Sayeed, Q. Liao, D. Faucher, E. Grinshpun, and S. Sharma, "Cloud analytics for wireless metric prediction—Framework and performance," in *Proc. IEEE Int. Conf. Cloud Comput. (CLOUD)*, New York, NY, USA, 2015, pp. 995–998.

[95] J. Tadrous, A. Eryilmaz, and H. El Gamal, "Proactive resource allocation: Harnessing the diversity and multicast gains," *IEEE Trans. Inf. Theory*, vol. 59, no. 8, pp. 4833–4854, Aug. 2013.

[96] L. Huang, S. Zhang, M. Chen, and X. Liu, "When backpressure meets predictive scheduling," in *Proc. ACM Int. Symp. Mobile Ad Hoc Netw. Comput. (MobiHoc)*, Philadelphia, PA, USA, 2014, pp. 33–42.

[97] N. Abedini and S. Shakkottai, "Content caching and scheduling in wireless networks with elastic and inelastic traffic," *IEEE/ACM Trans. Netw.*, vol. 22, no. 3, pp. 864–874, Jun. 2014.

[98] Q. Xu, S. Mehrotra, Z. Mao, and J. Li, "PROTEUS: Network performance forecast for real-time, interactive mobile applications," in *Proc. ACM Int. Conf. Mobile Syst. Appl. Services (MobiSys)*, Taipei, Taiwan, 2013, pp. 347–360.

[99] S. Samulevicius, T. B. Pedersen, and T. B. Sorensen, "MOST: Mobile broadband network optimization using planned spatio-temporal events," in *Proc. IEEE Veh. Technol. Conf. (VTC Spring)*, Glasgow, U.K., 2015, pp. 1–5.

[100] M.-F. R. Lee, F.-H. S. Chiu, H.-C. Huang, and C. Ivancsits, "Generalized predictive control in a wireless networked control system," *Hindawi Int. J. Distrib. Sensor Netw.*, vol. 2013, Dec. 2013, Art. no. 475730.

[101] A. Balachandran *et al.*, "Developing a predictive model of quality of experience for Internet video," in *Proc. ACM SIGCOMM*, Hong Kong, 2013, pp. 339–350.

[102] F. Beister and H. Karl, "Predicting mobile video inter-download times with hidden Markov models," in *Proc. IEEE Int. Conf. Wireless Mobile Comput. Netw. Commun. (WiMob)*, Larnaca, Cyprus, 2014, pp. 359–364.

[103] E. Pollakis and S. Stanczak, "Anticipatory networking for energy savings in 5G systems," in *Proc. VDE ITG-Fachbericht-WSA*, 2016, pp. 1–7.

[104] H. Yu, M. H. Cheung, L. Huang, and J. Huang, "Predictive delay-aware network selection in data offloading," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, Austin, TX, USA, 2014, pp. 1376–1381.

[105] H. Yu, M. H. Cheung, L. Huang, and J. Huang, "Power-delay tradeoff with predictive scheduling in integrated cellular and Wi-Fi networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 735–742, Apr. 2016.

[106] J. Du, C. Jiang, Y. Qian, Z. Han, and Y. Ren, "Traffic prediction based resource configuration in space-based systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, 2016, pp. 1–6.

[107] J. Du, C. Jiang, Y. Qian, Z. Han, and Y. Ren, "Resource allocation with video traffic prediction in cloud-based space systems," *IEEE Trans. Multimedia*, vol. 18, no. 5, pp. 820–830, May 2016.

[108] K. Papagiannaki, N. Taft, Z.-L. Zhang, and C. Diot, "Long-term forecasting of Internet backbone traffic: Observations and initial models," in *Proc. IEEE INFOCOM*, San Francisco, CA, USA, 2003, pp. 1178–1188.

[109] N. Sadek and A. Khotanzad, "Multi-scale high-speed network traffic prediction using k-factor Gegenbauer ARMA model," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 4. Paris, France, 2004, pp. 2148–2152.

[110] B. Zhou, D. He, Z. Sun, and W. H. Ng, "Network traffic modeling and prediction with ARIMA/GARCH," in *Proc. HET-NETs Conf.*, 2005, pp. 1–10.

[111] H. Abou-Zeid and H. S. Hassanein, "Predictive green wireless access: Exploiting mobility and application information," *IEEE Wireless Commun.*, vol. 20, no. 5, pp. 92–99, Oct. 2013.

[112] J. Yao, S. S. Kanhere, and M. Hassan, "Improving QoS in high-speed mobility using bandwidth maps," *IEEE Trans. Mobile Comput.*, vol. 11, no. 4, pp. 603–617, Apr. 2012.

[113] H. Riiser, P. Vigmostad, C. Griwodz, and P. Halvorsen, "Commute path bandwidth traces from 3G networks: Analysis and applications," in *Proc. ACM Multimedia Syst. Conf. (MMSys)*, Oslo, Norway, 2013, pp. 114–118.

[114] P. Millan *et al.*, "Tracking and predicting end-to-end quality in wireless community networks," in *Proc. IEEE Int. Conf. Future Internet Things Cloud (FiCloud)*, Rome, Italy, 2015, pp. 794–799.

[115] X. Yin, A. Jindal, V. Sekar, and B. Sinopoli, "A control-theoretic approach for dynamic adaptive video streaming over HTTP," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 45, no. 4, pp. 325–338, 2015.

[116] Y. Sun *et al.*, "CS2P: Improving video bitrate selection and adaptation with data-driven throughput prediction," in *Proc. ACM SIGCOMM*, Florianopolis, Brazil, 2016, pp. 272–285.

[117] J. Jiang *et al.*, "CFA: A practical prediction system for video QoE optimization," in *Proc. USENIX Symp. Netw. Syst. Design Implement. (NSDI)*, Santa Clara, CA, USA, 2016, pp. 137–150.

[118] A. H. Zahran *et al.*, "OSCAR: An optimized stall-cautious adaptive bitrate streaming algorithm for mobile networks," in *Proc. ACM Int. Workshop Mobile Video (MoVid)*, Klagenfurt, Austria, 2016, p. 2.

[119] C. Wang, A. Rizk, and M. Zink, "SQUAD: A spectrum-based quality adaptation for dynamic adaptive streaming over HTTP," in *Proc. ACM Int. Conf. Multimedia Syst. (MMSys)*, Klagenfurt, Austria, 2016, p. 1.

[120] K. Miller, D. Bethanabhotla, G. Caire, and A. Wolisz, "A control-theoretic approach to adaptive video streaming in dense wireless networks," *IEEE Trans. Multimedia*, vol. 17, no. 8, pp. 1309–1322, Aug. 2015.

[121] E. Baştuğ, J.-L. Guénégo, and M. Debbah, "Proactive small cell networks," in *Proc. IEEE Int. Conf. Telecommun. (ICT)*, 2013, pp. 1–5.

[122] E. Baştuğ, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5G wireless networks," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 82–89, Aug. 2014.

[123] E. Baştuğ, M. Bennis, and M. Debbah, "Anticipatory caching in small cell networks: A transfer learning approach," in *Proc. 1st KuVS Workshop Anticipatory Netw.*, Stuttgart, Germany, 2014, pp. 1–3.

[124] V. A. Siris, X. Vasilakos, and D. Dimopoulos, "Exploiting mobility prediction for mobility & popularity caching and dash adaptation," in *Proc. IEEE World Wireless Mobile Multimedia Netw. (WoWMoM)*, Coimbra, Portugal, 2016, pp. 1–8.

[125] N. Golrezaei, K. Shanmugam, A. G. Dimakis, A. F. Molisch, and G. Caire, "Femtocaching: Wireless video content delivery through distributed caching helpers," in *Proc. IEEE INFOCOM*, Orlando, FL, USA, 2012, pp. 1107–1115.

[126] J. Tadrous and A. Eryilmaz, "On optimal proactive caching for mobile networks with demand uncertainties," *IEEE/ACM Trans. Netw.*, vol. 24, no. 5, pp. 2715–2727, Oct. 2016.

[127] J. Tadrous, A. Eryilmaz, and H. El Gamal, "Joint smart pricing and proactive content caching for mobile services," *IEEE/ACM Trans. Netw.*, vol. 24, no. 4, pp. 2357–2371, Aug. 2016.

[128] Y. Gu, W. Saad, M. Bennis, M. Debbah, and Z. Han, "Matching theory for future wireless networks: Fundamentals and applications," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 52–59, May 2015.

[129] O. Semiari, W. Saad, S. Valentin, M. Bennis, and H. V. Poor, "Context-aware small cell networks: How social metrics improve wireless resource allocation," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 5927–5940, Nov. 2015.

[130] O. Semiari, W. Saad, and M. Bennis, "Context-aware scheduling of joint millimeter wave and microwave resources for dual-mode base stations," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, 2016, pp. 1–6.

[131] N. Namvar, W. Saad, B. Maham, and S. Valentin, "A context-aware matching game for user association in wireless small cell networks," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Florence, Italy, 2014, pp. 439–443.

[132] Y. Zhang *et al.*, "Social network aware device-to-device communication in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 177–190, Jan. 2015.

[133] K. Hamidouche, W. Saad, and M. Debbah, "Many-to-many matching games for proactive social-caching in wireless small cell networks," in *Proc. IEEE Int. Symp. Model. Optim. Mobile Ad Hoc Wireless Netw. (WiOpt)*, Hammamet, Tunisia, 2014, pp. 569–574.

[134] A. Noulas, S. Scellato, N. Lathia, and C. Mascolo, "Mining user mobility features for next place prediction in location-based services," in *Proc. IEEE Int. Conf. Data Min. (ICDM)*, Brussels, Belgium, 2012, pp. 1038–1043.

[135] F. Calabrese, G. D. Lorenzo, and C. Ratti, "Human mobility prediction based on individual and collective geographical preferences," in *Proc. IEEE Int. Conf. Intell. Transp. Syst. (ITSC)*, Funchal, Portugal, 2010, pp. 312–317.

[136] H. Bapierre, G. Groh, and S. Theiner, "A variable order Markov model approach for mobility prediction," *Pervasive Comput.*, pp. 8–16, 2011. AQ10

[137] M. Proebster, M. Kaschub, T. Werthmann, and S. Valentin, "Context-aware resource allocation for cellular wireless networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2012, no. 1, p. 216, 2012.

[138] M. Proebster, M. Kaschub, and S. Valentin, "Context-aware resource allocation to improve the quality of service of heterogeneous traffic," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kyoto, Japan, 2011, pp. 1–6.

[139] Z. Yi, X. Dong, X. Zhang, and W. Wang, "Spatial traffic prediction for wireless cellular system based on base stations social network," in *Proc. IEEE Syst. Conf. (SysCon)*, Orlando, FL, USA, 2016, pp. 1–5.

[140] G. I. Tsiropoulos, D. G. Stratogiannis, N. Mantas, and M. Louta, "The impact of social distance on utility based resource allocation in next generation networks," in *Proc. IEEE Int. Congr. Ultra Mod. Telecommun. Control Syst. Workshops (ICUMT)*, Budapest, Hungary, 2011, pp. 1–6.

[141] M. O. Jackson, *Social and Economic Networks*. Princeton, NJ, USA: Princeton Univ. Press, 2008.

[142] Telecom Italia. *Big Data Challenge 2015*. [Online]. Available: http://aris.me/contents/teaching/data-mining-2015/project/BigDataChallengeData.html

[143] A. C. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge, U.K.: Cambridge Univ. Press, 1990.

[144] Z. R. Zaidi and B. L. Mark, "Real-time mobility tracking algorithms for cellular networks based on Kalman filtering," *IEEE Trans. Mobile Comput.*, vol. 4, no. 2, pp. 195–208, Mar./Apr. 2005.

[145] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filtering theory," *Elsevier Transp. Res. B Methodol.*, vol. 18, no. 1, pp. 1–11, 1984.

[146] G. P. Pappas and M. A. Zohdy, "Extended Kalman filtering and pathloss modeling for shadow power parameter estimation in mobile wireless communications," *Int. J. Smart Sens. Intell. Syst.*, vol. 7, no. 2, pp. 898–924, 2014.

[147] J. Lee, M. Sun, and G. Lebanon, "A comparative study of collaborative filtering algorithms," *arXiv preprint arXiv:1205.3193*, 2012. AQ11

[148] E. Baştuğ, M. Bennis, and M. Debbah, *Think Before Reacting: Proactive Caching in 5G Small Cell Networks*. Wiley, 2015. AQ12

[149] S. Dutta, A. Narang, S. Bhattacherjee, A. S. Das, and D. Krishnaswamy, "Predictive caching framework for mobile wireless networks," in *Proc. IEEE Int. Conf. Mobile Data Manag. (MDM)*, Pittsburgh, PA, USA, 2015, pp. 179–184.

[150] R. Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Trans. Neural Netw.*, vol. 16, no. 3, pp. 645–678, May 2005.

[151] S. K. Murthy, "Automatic construction of decision trees from data: A multi-disciplinary survey," *Kluwer Data Min. Knowl. Disc.*, vol. 2, no. 4, pp. 345–389, 1998.

[152] J. O. Ramsay, *Functional Data Analysis*. Wiley, 2006.

[153] J. O. Ramsay and C. Dalzell, "Some tools for functional data analysis," *JSTOR J. Roy. Stat. Soc. B (Methodol.)*, vol. 53, no. 3, pp. 539–572, 1991.

[154] M. C. Mozer, R. Wolniewicz, D. B. Grimes, E. Johnson, and H. Kaushansky, "Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry," *IEEE Trans. Neural Netw.*, vol. 11, no. 3, pp. 690–696, May 2000.

[155] H. Kaaniche and F. Kamoun, "Mobility prediction in wireless ad hoc networks using neural networks," *J. Telecommun.*, vol. 2, no. 1, pp. 95–101, 2010.

[156] C. Chen, X. Zhu, G. de Veciana, A. C. Bovik, and R. W. Heath, "Rate adaptation and admission control for video transmission with subjective quality constraints," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 1, pp. 22–36, Feb. 2015.

[157] C. Chen, R. W. Heath, A. C. Bovik, and G. de Veciana, "A Markov decision model for adaptive scheduling of stored scalable videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 6, pp. 1081–1095, Jun. 2013.

[158] D. Bianchi, A. Ferrara, and M. D. Di Benedetto, "Networked model predictive traffic control with time varying optimization horizon: The Grenoble South Ring case study," in *Proc. IEEE Eur. Control Conf. (ECC)*, Zürich, Switzerland, 2013, pp. 4039–4044.

[159] K. Witheephanich, J. M. Escaño, D. M. de la Peña, and M. J. Hayes, "A min–max model predictive control approach to robust power management in ambulatory wireless sensor networks," *IEEE Syst. J.*, vol. 8, no. 4, pp. 1060–1073, Dec. 2014.

[160] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[161] A. Schrijver, *Theory of Linear and Integer Programming*. Chichester, U.K.: Wiley, 1998.

[162] S. J. Qin and T. A. Badgwell, "A survey of industrial model predictive control technology," *Elsevier Control Eng. Pract.*, vol. 11, no. 7, pp. 733–764, 2003.

[163] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York, NY, USA: Wiley, 2014.

[164] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, vol. 1. Cambridge, MA, USA: MIT Press, 1998.

[165] F. Fu and M. van der Schaar, "A systematic framework for dynamically optimizing multi-user wireless video transmission," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 3, pp. 308–320, Apr. 2010.

[166] E. Hossain and M. Hasan, "5G cellular: Key enabling technologies and research challenges," *IEEE Instrum. Meas. Mag.*, vol. 18, no. 3, pp. 11–21, Jun. 2015.

[167] S. Giordano *et al.*, "Mobile ad hoc networks," *Handbook of Wireless Networks and Mobile Computing*, 2002, pp. 325–346. AQ13

[168] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 1801–1819, 4th Quart., 2014.

[169] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2347–2376, 4th Quart., 2015.

[170] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for smart cities," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 22–32, Feb. 2014.

[171] L. D. Xu, W. He, and S. Li, "Internet of Things in industries: A survey," *IEEE Trans. Ind. Informat.*, vol. 10, no. 4, pp. 2233–2243, Nov. 2014.

[172] H. Zimmermann, "OSI reference model—The ISO model of architecture for open systems interconnection," *IEEE Trans. Commun.*, vol. 28, no. 4, pp. 425–432, Apr. 1980.

[173] NGMN. *Next Generation Mobile Networks*. [Online]. Available: http://www.ngmn.de/publications/all-downloads/article/ngmn-5g-white-paper.html

[174] I. Malanchini, S. Valentin, and O. Aydin, "Wireless resource sharing for multiple operators: Generalization, fairness, and the value of prediction," *Elsevier Comput. Netw.*, vol. 100, pp. 110–123, May 2016.

[175] G. P. Fettweis, "The tactile Internet: Applications and challenges," *IEEE Veh. Technol. Mag.*, vol. 9, no. 1, pp. 64–70, Mar. 2014.

[176] V. Suryaprakash and I. Malanchini, "Reliability in future radio access networks: From linguistic to quantitative definitions," in *Proc. IEEE/ACM Int. Symp. Qual. Service (IWQoS)*, Beijing, China, 2016, pp. 1–2.

[177] N. Singer, *Sharing Data, But Not Happily*, New York Times, New York, NY, USA, 2015. accessed on Nov. 5, 2016. [Online]. Available: http://www.nytimes.com/2015/06/05/technology/consumers-conflicted-over-data-mining-policies-report-finds.html?_r=0

[178] J. Wan, D. Zhang, S. Zhao, L. Yang, and J. Lloret, "Context-aware vehicular cyber-physical systems with cloud support: Architecture, challenges, and solutions," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 106–113, Aug. 2014.

Nicola Bui, photograph and biography not available at the time of publication.

Matteo Cesana, photograph and biography not available at the time of publication.

S. Amir Hosseini, photograph and biography not available at the time of publication.

Qi Liao, photograph and biography not available at the time of publication.

Ilaria Malanchini, photograph and biography not available at the time of publication.

Joerg Widmer, photograph and biography not available at the time of publication.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

PLEASE NOTE: We cannot accept new source files as corrections for your paper. If possible, please annotate the PDF proof we have sent you with your corrections and upload it via the Author Gateway. Alternatively, you may send us your corrections in list format. You may also upload revised graphics via the Author Gateway.

AQ1: Please be advised that per instructions from the Communications Society this proof was formatted in Times Roman font and therefore some of the fonts will appear different from the fonts in your originally submitted manuscript. For instance, the math calligraphy font may appear different due to usage of the usepackage[mathcal]euscript. The Communications Society has decided not to use Computer Modern fonts in their publications.
AQ2: Please confirm/give details of funding source.
AQ3: Please provide the postal code for "Politecnico di Milano, Milano, Italy."
AQ4: Note that if you require corrections/changes to tables or figures, you must supply the revised files, as these items are not edited for you.
AQ5: Please provide the in-text citation for Table V.
AQ6: Please confirm the volume number for References [8], [28], [32], [100], and [137].
AQ7: Please confirm if the location and publisher information for References [28] and [54] are correct as set.
AQ8: Please provide the technical report number for Reference [34].
AQ9: Please provide the accessed date for References [36], [142], and [174].
AQ10: Please provide the volume number and the issue number or month for Reference [136].
AQ11: Please provide the complete details and exact format for Reference [147].
AQ12: Please provide the location for References [148] and [152].
AQ13: Please provide the publisher name and location for Reference [167].