Pursuing the Potential of New Mechanisms for Performance Engineering of 5G

Nikolaos Pappas, Emmanouil Fountoulakis, Cristian Tatino, Vangelis Angelakis and Di Yuan

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-147968>

N.B.: When citing this work, cite the original publication.

Pappas, N., Fountoulakis, E., Tatino, C., Angelakis, V., Yuan, Di, (2017), Pursuing the Potential of New Mechanisms for Performance Engineering of 5G, *2017 IEEE 22ND INTERNATIONAL WORKSHOP ON COMPUTER AIDED MODELING AND DESIGN OF COMMUNICATION LINKS AND NETWORKS (CAMAD)*. <https://doi.org/10.1109/CAMAD.2017.8031631>

Original publication available at:

<https://doi.org/10.1109/CAMAD.2017.8031631>

Copyright: IEEE

<http://www.ieee.org/>

©2017 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Pursuing the Potential of New Mechanisms for Performance Engineering of 5G

Nikolaos Pappas*, Emmanouil Fountoulakis*, Cristian Tatino*, Vangelis Angelakis*, Di Yuan*

*Department of Science and Technology, Linköping University, Sweden

Email: {nikolaos.pappas, emmanouil.fountoulakis, cristian.tatino, vangelis.angelakis, di.yuan}@liu.se

Abstract—In this position paper, we provide a discussion of the need for new mechanisms towards optimizing the performance of the fifth generation (5G) wireless systems, and then we outline on-going research work and results in this direction. Initially we go over the recently proposed concept of network anticipation and then resource allocation, since these two themes form the core of the ACT5G research project. Then, we discuss two specific topics within the project. The first is the use of stochastic modeling for millimeter wave (mmWave) systems, which are expected to be part of 5G networks. Our modeling enables network anticipation, resulting in an estimation of the coverage probability of mmWave, including coverage with non-line-of-sight (NLOS) mmWave beams. In the second research topic, we address the use of scalable and dynamic transmission time interval (TTI) for scenarios with a mix of different traffic types. In addition, and highlight the importance of optimizing the channel and TTI allocation in this context. Numerical results of these studies are presented and discussed. Finally, we outline directions for future study.

I. INTRODUCTION

The rapid growth of data transferred over in mobile broadband network calls for sustainable capacity growth and performance improvements in the next generation of wireless systems, namely 5G. The success of 5G will heavily rely on delivering satisfactory user experience at a low operational cost. In this context, research and development (R&D) is required not only for technological innovation at the device level, but also management system design for the flexibility which is necessary to predict and adapt based on network usage. To this end, network anticipation is a promising approach. By predicting and adapting to coverage conditions and other performance indicators at various time scales, an anticipatory-enabled 5G network will dramatically improve the operation quality and efficiency in comparison to systems without this capability. This is of particular importance since new spectrum is considered in 5G. A good example is the use of millimeter

wave (mmWave) systems, for which the propagation and channel characteristics are quite different from the low frequency bands. A lack of tools for predicting the performance, such as coverage, hampers the successful deployment of mmWave to deliver its potential. The approach of anticipation and prediction will enable to manage optimally the spectrum and network resource, as well as to allow network operators to deliver service with reduced expenditure via sharing infrastructure and resources.

A second important direction in 5G is resource allocation. In order to achieve higher data rate and at the same time deal with mixed types traffic, e.g., services with small amount of data but stringent delay requirements, and throughput-oriented data service, both the physical and access layers require new solutions and design. At the same time, resource allocation, such as scheduling, has to take into account the new possibilities enabled by the new design solutions, as well as new types of deployment scenarios (i.e., very dense networks).

As the dimension and complexity of the mobile networks is scaling up, the complexity of prediction and resource allocation pose great challenges. To address the challenges, techniques for current networks are no longer sufficient. On the other hand, 4G/5G networks also have distinctive features with respect to legacy networks. First, the quality of the available network data is dramatically increasing as the network elements are becoming able to store more refined measurements. Second, the quantity of such data is also going to scale up due to densification of network deployments.

II. LITERATURE REVIEW

The literature has been mostly leveraging network data for estimation and diagnosis of faults in mobile networks. The tenet is to feedback key performance indicators (KPIs) from the network into the radio resource management logics. Available research can be classified with respect to (i) the anomalies detected and

(ii) the targeted network technology. Automatic tools focusing on LTE [3] and Self Organizing Networks have been considered to enhance the performance of under-performing radio cells. Bayes Classifiers have been a successful tool in classifying anomalies out of alarms coming from the radio access network elements [4]. Generally speaking, the work in the literature proposes tools for assessing/classifying the network performance offline by resorting to, typically, computationally heavy algorithms from machine learning [5].

Utilizing this bulk of data has only recently drawn the attention of the research community to generate predictions on how the network conditions will evolve. In a proactive network [6], one considers that the prediction of traffic can be catered for resource allocation, treating the network as a delay tolerant network (DTN) with hard deadlines. In a similar setup, the anticipation of requested service rates for video streaming drives a scheduling scheme to avoid buffer under-runs [7]. Mobility estimation patterns have been also exploited for vehicular users to provide a capacity estimator for relatively short time-scales [8].

For resource allocation, there has been a large amount of literature on various resource optimization problems for orthogonal frequency division multiple access (OFDMA), for which the assignment of sub-carriers and power can both be subject to optimization. See, for example, [18]–[22]. At the network level, one approach for linking the performance to network configuration is the use of the so called load-coupling model which applies interference averaging to model mutual interference (e.g., [23]–[25]). This approach allows for a reasonable approximation for performance prediction, with the advantage of tractability.

More recent works on resource allocation have started looking into more advanced access scheme. One example is the non-orthogonal multiple access (NOMA) that deployed multi-user detection and interference cancellation, such that the same time-frequency resource unit can be reused by multiple users. Optimization models and algorithms for resource allocation in NOMA can be found in, for example, [26]–[30].

III. RESEARCH OVERVIEW

Currently, a number of active research projects, are putting efforts on prediction and resource allocation for 5G [31]. In the anticipatory networking techniques in 5G and beyond (ACT5G) project, our research addresses the challenges in anticipatory networking for next generation

wireless systems. As the general objectives, the ongoing research aims to:

- 1) investigate analytic models and methods to effectively and efficiently evaluate the performance at various time scales, addressing the trade-off of accuracy versus complexity,
- 2) develop network resource and management optimization techniques and algorithms that are tailored for new physical layer structures and access schemes proposed for 5G networks.

The above two objectives fall into the themes of anticipation and reaction techniques, respectively. These two themes form the core of the ACT5G research project. The themes are ultimately intertwined, as efficient resource allocation shall leverage the power of anticipation and prediction techniques, whereas the latter has little value if not put in the context of performance optimization.

A. Approach

ACT5G focuses on models, methods, concepts, and algorithms for network anticipation and network reaction. The approach taken for performance prediction and resource allocation consists in analytical modeling and numerical optimization. For modeling, tools from a number of fields, such as stochastic geometry, learning theory, and compressive sensing are of relevance. The optimization techniques to be adopted will primarily be based on the development of fast, low-complexity algorithms for online use. At the same time, the project develops methods that approach the global optimum for benchmarking.

B. Expected Outcomes and Contributions

Overall, research carried out within the ACT5G project is expected to advance the state of the art in efficiently performing network prediction and adaptive resource allocation. The main contributions consist of

- 1) new analytical tools and methods for estimating and predicting network performance and data-analysis algorithms for network status estimation addressing timescale, accuracy, and complexity;
- 2) algorithms for link outage and delay estimation, and user rate prediction for effective prefetching and buffering.
- 3) deriving computationally efficient algorithms that leverage the power of prediction for flexible and efficient resource allocation with the presence of heterogeneous service requirements.

In the forthcoming two sections, we illustrate the ongoing research of ACT5G by presenting some recent work in analytical modeling of mmWave systems for coverage prediction, and optimal scheduling and resource allocation for scalable transmission time interval (TTI).

IV. STOCHASTIC MODELING FOR MMWAVE SYSTEMS

One of the possible means to high rate data in 5G is to consider mmWave frequency bands from 30 to 300 GHz. At such high frequencies, the main challenges originate from large path loss and link blockage effects. The focus of current research of mmWave communications is the study of propagation characteristics, channel modeling, beam forming and medium access control design. Here, we develop a stochastic model for characterizing the coverage probability of a beam in outdoor mmWave systems [17]. Whereas it is expected that the coverage is mainly provided by a beam forms directly toward the user, the reflections can be exploited in non-line-of-sight (NLOS) conditions to overcome the blockage effect of the obstacles that are modeled using stochastic processes. To our knowledge, there is analytical analysis of the coverage probability using stochastic models taking into account both line-of-sight (LOS) and potential non-LOS (NLOS) links. This analysis provides an interesting tool to investigate the effects of user mobility and beam selection in mmWave cellular systems. The main novelty lies in the coupling of the following two aspects.

- The model incorporates beamforming, hence it allows to evaluate the coverage probability for any given beam, by taking into account the beam features (i.e., orientation and width) as well as the transmitter-receiver position.
- The model evaluates the coverage probability not only considering the direct beam but also including first-order reflections, which contribute to the coverage probability in NLOS conditions.

TABLE I: Coverage probability (in %) with respect to distance to base station.

Distance	Direct beam	Direct and reflected beams
25	0.50	0.55
50	0.38	0.46
75	0.28	0.39
100	0.21	0.33

The numerical experiments show that the proposed model is able to capture the dependency of the beam coverage probability on various parameters, such as user position, beam orientation and width, and obstacle density. In Table I, we provide sample results of the coverage probability with respect to the user distance to the base station. One can observe that although the major portion of the coverage probability is contributed by the beam that covers the user directly, reflections can still contribute to enhancing coverage, especially for larger user distances, i.e., when the line-of-sight (LOS) probability dramatically decreases. Indeed, for large distance, the coverage due to reflection can be considerable.

TABLE II: Comparison between the analytical model and simulation.

Distance	Analytical		Simulation	
	Direct	Reflected	Direct	Reflected
25	0.50	0.06	0.49	0.06
50	0.38	0.06	0.38	0.06
75	0.28	0.06	0.27	0.06
100	0.21	0.06	0.20	0.05

The sample results in Table II compare the analytical model with the simulation results. Here, the beam width is 10° . It can be observed that the model provides good accuracy in providing the coverage probability, and hence it is very useful for predicting coverage of mmWave systems. We remark that when the beam width increases, the gap between analytical model and simulation grows, though the accuracy remains reasonable.

V. SCALABLE TTI

In order to pursue higher rate, lower delay, and more flexibility in resource allocation, the concept of scalable TTI has been recently introduced [32], [33]. The basic idea is to dynamically adapt the length of the TTI to the service requirements. For example, the so called mission critical communications (MCC) are characterized by small packets and the requirement for ultra low latency of less than 1 ms and high reliability. Adopting a very short TTI would fulfill the need of MCC services. However, shorter TTI comes with the cost of additional signaling overhead and hence waste of resource for the conventional mobile broadband (MBB) services.

To meet the requirements of mixed services types, scalable TTI allows the TTI duration to be changed for each scheduling instance. In [16], we have analyzed the potential benefits of using dynamic TTI length. We

proposed a scheduling approach that optimizes the TTI length and channel allocation according to the services deadlines, the amount of data to be delivered for each service, and the conditions on the individual channels. The system works on an instance-by-instance basis as a practical scheduler. That is, once a scheduling decision is made, the deadline and the remaining amount of data are updated before the next scheduling instance, for which the channel conditions are not known when deriving the current scheduling solution. Moreover, new arrivals may occur. The objective is a utility function that encourages to allocate resource to services with short deadline or to services achieving high emptying rate. However, the utility function alone cannot guarantee that the scheduler allocates bandwidth to the services which can be completely served. Thus, we introduce a function that increases as the number of served services increases. Therefore, the scheduler caters as many services as possible at each scheduling time while striving to fulfill the deadline requirements of the services. For the resulting scalable-TTI enabled channel allocation (STCA) problem, the following contributions are made.

- We proved that STCA is NP-hard.
- If for each service, the channels are flat, STCA was proved to be tractable using a polynomial-time dynamic programming algorithm.
- We developed an Integer Linear Program (ILP) that finds the optimal solution (OS) for each scheduling instance for the given utilization function.
- A suboptimal but fast and hence practical channel allocation algorithm has been derived.

For the simulation setup, we consider three MCC sources and one MBB source that generate services which require to be served. We provide results for variable arrival rate of MCC services and fixed arrival rate for MBB services. Some sample performance results are shown in Tables III-IV. In both tables, the first column refers to the normalized arrival intensity of MCC service. For fixed TTI, the time length is 1 ms. The set of possible TTI lengths for scalable TTI is $[0.1, 0.2, \dots, 1]$ ms. The results for both fixed and scalable TTI are obtained using the optimal solution for each scheduling instance. The last column refers to the shortest deadline first scheduling (SDFS), in which the services are allocated resource only according to their deadlines.

The results in Tables III-IV show, respectively, the percentage satisfied MCC service and the throughput of MBB service.

From the simulation results, we observe that the STCA

TABLE III: Performance of scalable TTI for MCC service.

MCC intensity	Scalable TTI	Fixed TTI	SDFS
1.0	95%	98%	87%
1.5	93%	96%	77%
2.0	78%	81%	53%

TABLE IV: Performance of scalable TTI for MBB throughput (Mbps).

MCC intensity	Scalable TTI	Fixed TTI	SDFS
1.0	9.2	7.1	1.4
1.5	4.6	2.7	0.4
2.0	3.3	0.9	0.1

scheduler (using either fixed or scalable TTI) outperforms SDFS, as the latter only emphasize on deadline, no matter if a service will be completely served or not. For the sample results, fixed TTI is able to meet slightly more MCC services than scalable TTI. This is a consequence of the fact that the scheduler with scalable TTI strives to serve as many services as possible by selecting longer TTI lengths in general than the fixed TTI. Therefore, when the amount of incoming MCC services increases, the scheduler has to cope with services whose deadlines get closer. However, we observe that STCA with scalable TTI considerably outperforms the fixed TTI for the MBB services. The STCA scheduler with scalable TTI length selects the channel allocation and the TTI length by taking into account both the deadline constraints of the MCC services and the throughput requirements of the MBB services. Hence, the STCA scheduler with scalable TTI length provides a good trade-off between the amount of served MCC services and the throughput of MBB services.

VI. CONCLUSIONS AND FUTURE WORK

The evolution toward 5G networks calls for a host of new techniques and concepts. In this paper, we have outlined our ongoing efforts along this line of research, within the ACT5G project and we presented two specific topics under investigation: stochastic modeling for mmWave systems, and resource allocation with scalable TTI. As conclusions, the type of modeling approach for mmWave opens new opportunities for performance characterization and evaluation, and scalable TTI is promising, provided that engineering methods and optimization concepts are developed to adapt resource allocation with this concept.

There are numerous interesting topics for further investigation. One topic is to plug-in the system modeling approach for mmWave in resource optimization, e.g., dynamic beam design and beam power allocation. For scalable TTI, the next step consists in two-dimensional resource allocation, where the TTI size may vary by user in the same scheduling instance, and to pursue the potential enabled by this flexibility.

ACKNOWLEDGMENT

The work discussed and outlined in this paper is supported by European Union Horizon 2020 MSCA project ACT5G (contract number 643002).

REFERENCES

- [1] V. Angelakis, S. Fowler et al. Evaluation and Prospects from a Measurement Campaign on Real Multimedia Traffic in LTE vs. UMTS, invited paper, to appear in IEEE GWS-WVITAE, 2014.
- [2] M. Tiwana, B. Sayrac, et al. "Troubleshooting of 3GPP LTE mobility parameters using iterative statistical model refinement," in Proc. IFIP Wireless Days, 2009.
- [3] R. Barco, P. Lazaro, and P. Munoz. "A unified framework for self-healing in wireless networks," IEEE Communications Magazine, 50(12), 2012.
- [4] R. M. Khanafer, et al. "Automated diagnosis for UMTS networks using Bayesian network approach," IEEE Trans. on Vehicular Technology 57(4), 2008.
- [5] B. Bojovic, et al. "A study on machine learning and regression based models for performance estimation of LTE HetNets," 3rd ACROPOLIS Annual Workshop, 2013.
- [6] J. Tardus, A. Eryilmaz and H. El Gamal. "Proactive resource allocation: harnessing the diversity and multicast gains," IEEE Transactions on Information Theory, 59(8), 2013.
- [7] M. Draxler, S. Valentin, et al. "Anticipatory buffer control and quality selection for wireless video streaming," in arXiv:1309.5491 [cs.NI], 2013.
- [8] Z. Lu and G. De Vecianna. "Optimizing stored video delivery for mobile networks: the value of knowing the future," in Proc. IEEE INFOCOM, 2013.
- [9] M. Firooz and S. Roy, "Network tomography via compressed sensing," in Proc. IEEE GLOBECOM, 2010.
- [10] T. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. Wong, J. Schulz, M. Samimi, and F. Gutierrez. "Millimeter wave mobile communications for 5G cellular: It will work!" IEEE Access, vol. 1, pp. 335-349, 2013.
- [11] S. Rangan, T. Rappaport, and E. Erkip. "Millimetre-wave cellular wireless networks: Potentials and challenges," Proceedings of the IEEE, vol. 102, no. 3, pp. 366-385, March 2014.
- [12] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar. "Millimetre-wave beamforming as an enabling technology for 5g cellular communications: Theoretical feasibility and prototype results," IEEE Communications Magazine, vol. 52, no. 2, pp. 106-113, February 2014.
- [13] T. Bai, R. Vaze and R. W. Heath. Analysis of blockage effects on Urban cellular networks". IEEE Transactions on Wireless Communications, vol. 13, no. 9, pp. 5070-5083, Sept. 2014.
- [14] N. A. Muhammad, P. Wang, Y. Li, and B. Vucetic. "Analytical model for outdoor millimetre wave channels using geometry-based stochastic approach," IEEE Transactions on Vehicular Technology, vol. 66, no.2, pp.912-926, 2017.
- [15] M. R. Akdeniz et al. "Millimetre wave channel modeling and cellular capacity evaluation," IEEE Journal on Selected Areas in Communications, vol. 32, no. 6, pp. 1164-1179, June 2014.
- [16] E. Fountoulakis, N. Pappas, Q. Liao, V. Suryaprakash, and D. Yuan. "An examination of the benefits of scalable TTI for heterogeneous traffic management in 5G networks," RAWNET Workshop at WiOpt, May. 2017.
- [17] C. Tatino, I. Malanchini, D. Aziz, and D. Yuan. "Beam based stochastic model of the coverage probability in 5G millimeter wave systems," SPASWIN Workshop at WiOpt, May. 2017.
- [18] M. Bohge, J. Gross, A. Wolisz, and M. Meyer, "Dynamic resource allocation in OFDM systems: An overview of cross-layer optimization principles and techniques," IEEE Network Mag., vol. 21, no 1, pp. 53-59, Jan. 2007.
- [19] A. Feiten, R. Mathar, and M. Reyer. "Rate and power allocation for multiuser OFDM: an effective heuristic verified by branch-and-bound," IEEE Trans. Wireless Commun., vol. 7, no.1, pp. 60-64, Jan. 2008.
- [20] S. Sadr, A. Anpalagan, and K. Raahemifar, "Radio resource allocation algorithms for the downlink of multiuser OFDM communication systems," IEEE Commun. Surveys & Tutorials, vol. 11, no. 3, pp. 92-106, Third Quarter 2009.
- [21] T. Liu, C. Yang, and L.-L. Yang, "A low-complexity subcarrier-power allocation scheme for frequency-division multiple-access systems," IEEE Trans. Wireless Commun., vol. 5, no. 5, pp. 1571-1576, May 2010.
- [22] K. Seong, M. Mohseni, and J. M. Cioffi, "Optimal resource allocation for OFDMA downlink systems," in Proc. IEEE International Symposium on Infor. Theory (ISIT), Seattle, USA, Sep. 2006. pp. 1394-1398.
- [23] L. Lei, D. Yuan, C. K. Ho, and S. Sun, "Optimal cell clustering and activation for energy saving in load-coupled wireless networks," IEEE Transactions on Wireless Communications, vol. 14, no. 11, pp. 6150- 6163, 2015.
- [24] I. Siomina and D. Yuan, "Analysis of cell load coupling for LTE network planning and optimization," IEEE Transactions on Wireless Communications, vol. 11, no. 6, pp. 2287-2297, 2012.
- [25] A. J. Fehske, I. Viering, J. Voigt, C. Sartori, S. Redana, and G. P. Fettweis, "Small-cell self-organizing wireless networks," Proceedings of the IEEE, vol. 102, no. 3, pp. 334-350, 2014.
- [26] L. Dai, B. Wang, Y. Yuan, S. Han, C.-L. I, and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," IEEE Communications Magazine, vol. 53, no. 9, pp. 74-81, 2015.
- [27] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of nonorthogonal multiple access in 5G systems with randomly deployed users," IEEE Signal Processing Letters, vol. 21, no. 12, pp. 1501-1505, 2014.
- [28] J. Choi, "On the power allocation for a practical multiuser superposition scheme in NOMA systems," IEEE Communications Letters, vol. 20, no. 3, pp. 438-441, 2016.
- [29] S. Timotheou and I. Krikidis, "Fairness for non-orthogonal multiple access in 5G systems," IEEE Signal Processing Letters, vol. 22, no. 10, pp. 1647-1651, 2015.
- [30] L. Lei, D. Yuan, C. K. Ho, and S. Sun, "Power and channel allocation for non-orthogonal multiple access in 5G systems: Tractability and computation," IEEE Transactions on Wireless Communications, vol. 15, pp. 8580-8594, 2016.
- [31] N. Bui, M. Cesana, S. A. Hosseini, Q. Liao, I. Malanchini, J. Widmer, "A Survey of Anticipatory Mobile Networking: Context-Based Classification, Prediction Methodologies, and Optimization Techniques" IEEE Communications Surveys Tutorials, vol. PP, pp. 1-1, 2017.

- [32] Q. Liao, P. Baracca, D. Lopez-Perez, and L. G. Giordano, "Resource scheduling for mixed traffic types with scalable TTI in dynamic TDD systems," *IEEE Globecom Workshops*, 2016.
- [33] K. Pedersen, F. Frederiksen, G. Berardinelli, and P. Mogensen, "A flexible frame structure for 5G wide area," *IEEE 82nd Vehicular Technology Conference*, 2015.