5G Terminals with Multi-Streaming Features for Real-Time Mobile Broadband Applications

Tomislav SHUMINOSKI, Toni JANEVSKI

Ss. Cyril and Methodius University, Faculty of Electrical Engineering and Information Technologies,
Karpos 2 bb, 1000 Skopje, Republic of Macedonia

{tomish, tonij}@feit.ukim.edu.mk

Abstract. *In this paper we present a novel QoS framework on the network layer for 5G terminals with vertical multi-homing and multi-streaming capabilities by using radio networks aggregation. The proposed framework is leading to high performance utility networks with QoS provisioning for real-time multimedia services by achieving low packet delays, stochastic queuing network stability and highest mobile broadband capabilities i.e. bitrates. The proposed QoS algorithm is implemented within the mobile terminals on one side, and in dedicated proxy servers on mobile core network side. It is based on Lyapunov optimization techniques and it is targeted to handle simultaneously multiple multimedia service flows via multiple radio network interfaces in parallel.*

Keywords

5G, Lyapunov optimization, Quality of Service, Radio Access Technology, vertical multi-homing

1. Introduction

The exponential expansions and developments of mobile broadband networks are followed and supported by a novel research and development works towards the 5G era. This paper provides 5G framework that could lead to high performance delay-sensitive utility networks, spectral efficiency, network stability and high QoS provisioning for the key QoS parameters for any given multimedia service in nowadays and future Radio Access Technologies (RATs). Our framework is following the 5G key drivers and 5G enormous spectrum of advanced capabilities introduced in [1−7]. In that context our proposed algorithm with Lyapunov optimization technique implemented in both: the mobile terminal (MT) and a proxy server (placed in the core part of the mobile network) is following the 5G mobile networks' device-centric concept, with the multi-RAT or cloud-RAT paradigm using simultaneously multiple RAT interfaces. In order to exploit future 5G mobile broadband multi-RAT networks, and reaching the maximum average throughput, the proposed 5G terminal (5GT) is having vertical multi-homing and multi-streaming features [8], [9].

2. Related Work

The analysis and design methodology of our proposed 5G framework is based on the adaptive queuing Lyapunov optimization techniques [10–28], which are very powerful techniques for optimizing time average queuing networks and giving joint stability, efficiency and performance optimization. The existing solutions and applications are including maximization of the network uplink throughput subject to average power constraints on the terminal interfaces, minimizing average queue backlogs, subject to minimal queue network delay and in the same time achieving network stability and spectrum efficiency. Such Lyapunov drift-plus-penalty techniques were firstly applied to wireless networks in [12] by Tassiulas and Ephremides, where stochastic Lyapunov drift is used to develop a joint optimal routing and scheduling algorithm for the telecommunication networks. However, the Lyapunov drift has since become a powerful technique for the development of stable scheduling strategies for wireless and mobile systems and networks [15], [16], [19–24], [26–28], computer networks and switches [14], ad-hoc mobile networks [17], wireless sensor networks [18] and wireless mesh networks [25]. Moreover, example of services including maximization of the average network throughput with subject to average power constraints, minimizing the average power expenditure subject to network stability is presented in [15], [19], [22], [25] and [29]. In comparison with all above mentioned works, this paper applies a version of the Lyapunov drift-plus-penalty technique in the future 5G mobile and fixed terminals, here called with one name 5GTs, achieving high performance utility networks, spectral efficiency, network stability and high level of QoS provisioning for real-time mobile broadband services. Moreover, all changes and improvements we are implementing in the network layer of the 5GT, because the common part for all of existing and future RATs, which connect them all as a unifying technology, is IP.

Our proposed 5GT system model is device-centric,

multi-RAT, with several (*m*) interfaces; each for different RAT (overall M RATs) and has been previously presented in Fig. 1 [30]. However, despite the presented system model and AQUAplus optimization algorithm given in [30], the novelty of our framework presented in this paper is the development of novel adaptive QoS Module with advanced QoS user-centric network aggregation algorithm using vertical multi-homing and multi-streaming features for delay sensitive, real-time broadband services (applications). We refer to it as Massive throughput Advanced QoS-based User-centric Aggregation (MassiveAQUA) algorithm, which is defined independently from the wireless and mobile technologies and is doing Lyapunov drift-plus-penalty optimization for packet optimal scheduling in the 5GT with multiple RAT interfaces (in order to achieve massive throughput), achieving queue stability and simultaneously minimizing the queue delays, consequently having minimal overall end-to-end delay. Despite all, the AQUAplus algorithm has no mechanisms for dropping packets in the real queues, which leading to higher queue delays in comparison with the furthermore proposed MassiveAQUA. Plus, the MassiveAQUA algorithm is another extension of the AQUA algorithm (presented in related papers: [31] and [32]) placed within 5GMTs and proxy servers, because the AQUA algorithm is doing only multi-homing (selecting the best RAT for given service) with optimization problem solved by linear programming or genetic algorithms, without considering the network delay, the network stability and the queue backlogs.

The 5GT system model (Fig. 1) is almost the same as the system model in [30], with several differences: different utility functions (enriched with the delay constraints and penalties), different Lyapunov function, and existence of a packet drop decision vector for the network of queues – leading to delay sensitive and optimal multi-streaming and multi-homing processes. All novelties and differences with the AQUAplus algorithm optimization process are given in the following sections.

3. Optimization Algorithm

In Fig. 1 the 5GT system model with several (*m*) interfaces, each for different RAT (overall M RATs) is presented. Each source arrival process $x_i(t)$ after passing over the transport layer, is entering in the dedicated Network Layer Processor NLP$_i$ inside the Processors part of the Upper Network layer. Despite the real *M* queue network in [30], here we are considering *M* queue network with queue vector $Q(t) = (Q_1(t), Q_2(t), …, Q_M(t))$ that evolves in slotted time $t \in \{0,1,2,3,…\}$ with the following update equation:

$$Q_m(t+1) = \max[Q_m(t) + A_m(t) - \mu_m(t) - d_m(t), 0] \quad (1)$$

where $A_m(t) = \sum_{i=1}^{N} w_{i,m} x_i(t)$ is arrival rate variable for the

m-th queue, and the parameter $\mu_m(t)$ is output serving rate variable of the *m*-th queue, where $m \in \{1,2,3,…, M\}$. After
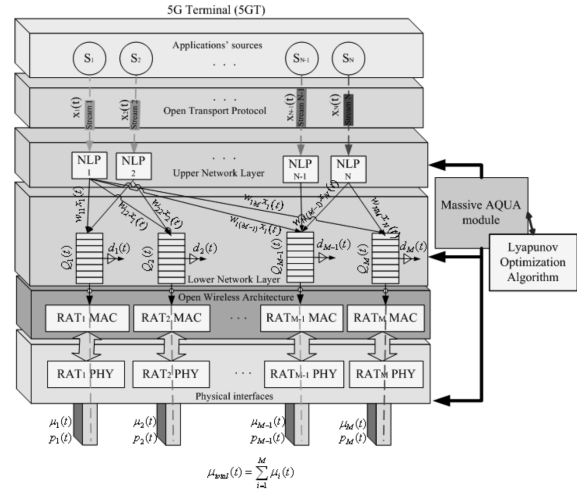

Fig. 1. The system model of the 5G terminal (fixed and mobile).

each queue we are trying to achieve maximal output serving rate $\mu_j(t)$ on each RAT interface $j = 1,2,.., M$, so the sum of all output serving rates, over the time, it will be also with maximal value. For the network of queues we are introducing a packet drop decision vector $d(t) = (d_1(t), d_2(t), . . . , d_M(t))$ and their adequate time average values: $\overline{d} \triangleq (\overline{d_1(t)}, \overline{d_2(t)}, ..., \overline{d_M(t)})$. The packet drop decision $d_m(t)$, of the *m*-th queue, allows packets already admitted to the network layer queue $Q_m(t)$ to be dropped if their delay is too large. Drop decisions $d_m(t)$ are chosen subject to the constraints:

$$0 \leq d_m(t) \leq A_{\max}, \text{ for } m \in \{1,2,3...M\} \quad (2)$$

where $A_{\max}$ is a finite constant, chosen to be the maximal value of all arrival rate variables $A_m(t)$ at a given time slot *t* (which limits the amount of traffic we can send into the Lower Network Layer). Furthermore, let us consider a random queue network state: $\psi(t) \triangleq [(\mu_1(t),..., \mu_M(t)); (A_1(t),...A_M(t))]$, and a control policy action $\alpha_{NL}(t)$ which is made every slot *t* with knowledge of the current state process $\psi(t)$ and is chosen within some abstract action set $A_{\psi(t)}$ (set containing all possible control actions). For simplicity we are assuming that the $\psi(t)$ process is i.i.d. over each slot.

Also, we are considering a vector with time average power values for each interface as $\overline{p} \triangleq (\overline{p_1(t)}, \overline{p_2(t)}, ..., \overline{p_M(t)})$ (the same as in [30]). Let the $\overline{x} \triangleq (\overline{x_1(t)}, \overline{x_2(t)}, ..., \overline{x_N(t)})$ be a vector of the time average arrival rates $(x_i(t))$. We are considering the following two separable utility functions of the above defined vectors:

$$\theta(\overline{x}) \triangleq \sum_{m=1}^{M} \sum_{i=1}^{N} w_{i,m} \overline{x_i(t)}, \text{ and} \quad (3)$$

$$\Gamma(\overline{p}, \overline{d}) \triangleq \sum_{m=1}^{M} \left(\log\left(1 + \frac{\overline{p_m(t)}}{n_m}\right) - \beta \overline{d_m(t)} \right) \quad (4)$$

where $\log(\cdot)$ denotes the natural logarithm, and the func-

tion $\log\left(1+\dfrac{\overline{p_m(t)}}{n_m}\right)$ has maximum derivative $1/n_m$, and is 0 when $p_m = 0$. The value n_m is finite constant and we assume $1/n_m < \infty$. Moreover, β is a constant that satisfies $1 \le \beta < \infty$. The utility function (3) is defined in [30] and equation (4) is novel utility function which incorporates the average packet drop decision vectors and the average capacity per radio interface, expressed with the logarithmic function given above. Then, for each 5GT system we are applying stochastic utility maximization framework to our streaming based system model, and the desired solution to the following optimization problem in uplink (when MT is sending information) and downlink (when the proxy server is sending information) is:

$$\text{Maximize: } \theta(\overline{\mathbf{x}}) + \Gamma(\overline{\mathbf{p}}, \overline{\mathbf{d}}) \qquad (5)$$

Subject to: 1. Time average stream (flow) over the queue $Q_m(t)$ is less or equal to the time average maximal output serving rate $\mu_m^{\max}$ on the interface m, i.e.:

$$\sum_{i=1}^{N} w_{i,m}\overline{x_i(t)} \le \mu_m^{\max} \; , \; \forall\, m \in \{1,2,...,M\} \qquad (6)$$

2. All queues $Q_m(t)$ are rate *stable*, i.e.:

$$\lim_{t\to\infty}\frac{\mathbf{E}\{|\,Q_m(t)\,|\}}{t} = 0 \,, \forall\, m \in \{1,2,...,M\} \qquad (7)$$

3. The desired time average power constraints are fulfilled [22] ($p_m(t)$ is the power incurred in interface m of the network on slot t, and p_m^{av} is a required time average power expenditure), i.e.:

$$\limsup_{t\to\infty}\overline{p_m(t)} \le p_m^{\mathrm{av}}, \forall\, m \in \{1,2,...,M\} \qquad (8)$$

4. For the control policy action: $\alpha_{\mathrm{NL}}(t) \in A_{\psi(t)}$.

5. For finite parameter $\mu_{\min} > 0$, the average output serving rate of the m-th queue satisfies:

$$\overline{\mu_m(t)} \ge \mu_{\min} \; , \; \forall\, m \in \{1,2,3...M\} \qquad (9)$$

The constraint (9) requires each queue $Q_m(t)$ to transmit with a time-average rate of at least $\mu_{\min}$, which ensures all queues are getting at least a minimum rate $\mu_{\min}$ of service. However, we shall not enforce this constraint, because if the input rate for the m-th queue is less than $\mu_{\min}$, then this constraint is wasteful.

For $t > 0$, the used variables as time average over the first t slots for $m \in \{1,2,3,...,M\}$ are defined with:

$$\overline{x_m(t)} \triangleq \lim_{t\to\infty}\frac{1}{t}\sum_{\tau=0}^{t-1}\mathbf{E}\{x_m(t)\}; \quad \overline{p_m(t)} \triangleq \lim_{t\to\infty}\frac{1}{t}\sum_{\tau=0}^{t-1}\mathbf{E}\{p_m(t)\};$$

$$\overline{\mu_m(t)} \triangleq \lim_{t\to\infty}\frac{1}{t}\sum_{\tau=0}^{t-1}\mathbf{E}\{\mu_m(t)\}; \quad \overline{d_m(t)} \triangleq \lim_{t\to\infty}\frac{1}{t}\sum_{\tau=0}^{t-1}\mathbf{E}\{d_m(t)\}. \qquad (10)$$

4. MassiveAQUA Algorithm

For solving the above optimization problem we are using the Lyapunov drift-plus-penalty method (using fixed positive penalty control parameters V_1 and V_2) given in [10]. We enforce the constraints 1 and 2 of the optimization problem with the actual queue $Q_m(t)$ (1) and the constraint 3 with virtual queue $Z_m(t)$ for each m, given in [30]. It is very easy to prove that the queues are mean rate stable [10] (because they are with finite queue length).

Also, to ensure worst-case delay is bounded and to enforce the constraint (9), we define a $\mu_{\min}$-persistent service queue, being a virtual queue $U_m(t)$ for each $m \in \{1,...,M\}$ with $U_m(0) = 0$ and with update equation:

$$U_m(t+1) = \max[U_m(t) - \mu_m(t) - d_m(t) + \mu_{\min}]$$
$$\text{if} \quad Q_m(t) > \mu_m(t) + d_m(t), \qquad (11)$$
$$\text{and} \quad U_m(t+1) = 0 \quad \text{if} \quad Q_m(t) \le \mu_m(t) + d_m(t).$$

We assume average throughout that $\mu_{\min} \le A_{\max}$. The condition $Q_m(t) \le \mu_m(t) + d_m(t)$ is satisfied whenever the backlog $Q_m(t)$ is cleared (by service and/or drops) on slot t. If this constraint is not active, then $U_m(t)$ has a departure process that is the same as $Q_m(t)$, but it has an arrival of size $\mu_{\min}$ every slot t. Furthermore, we are defining a combining queue vector $\mathbf{S}(t) \triangleq [\mathbf{Q}(t), \mathbf{Z}(t), \mathbf{U}(t)]$ and the Lyapunov function (see Sec. 3.1.2 for Lyapunov functions in [10]) as:

$$L(\mathbf{S}(t)) \triangleq \frac{1}{2}\left[\sum_{i=1}^{M}\left(Q_i^{\,2}(t) + Z_i^{\,2}(t) + U_i^{\,2}(t)\right)\right]. \qquad (12)$$

If we define the one-step conditional Lyapunov drift with $\Delta(\mathbf{S}(t)) \triangleq \mathbf{E}\{L(\mathbf{S}(t+1)) - L(\mathbf{S}(t))|\mathbf{S}(t)\}$, then the Lyapunov drift-plus-penalty expression is defined as:

$$\Delta(\mathbf{S}(t)) + V_1 \cdot \mathbf{E}\{\mathbf{u}(\mathbf{x}(t), \alpha_{\mathrm{NL}}(t), t)|\mathbf{S}(t)\} + $$
$$+ V_2 \cdot \mathbf{E}\{g(\mathbf{p}(t), \mathbf{d}(t), \alpha_{\mathrm{NL}}(t), t)|\mathbf{S}(t)\}. \qquad (13)$$

So, our proposed MassiveAQUA algorithm seeks to minimize the upper bound of (13) for all t, all possible values of $\mathbf{S}(t)$, and all control parameters $V_1 \ge 0$ and $V_2 \ge 0$. In (13) we are introducing a "penalty" vector process with $\mathbf{u}(\mathbf{x}(t),\,\alpha_{\mathrm{NL}}(t), t) = (u_0^1(t),\, u_0^2(t),..., u_0^M(t))$, which is already defined in more details in [30]. The novel "penalty" process $g(\mathbf{p}(t),\, \mathbf{d}(t),\, \alpha_{NL}(t),\, t)$, given here is defined with:

$$g(\mathbf{p}(t), \mathbf{d}(t), \alpha_{NL}(t), t) = $$
$$-\sum_{m=1}^{M}\log\left(1 + \frac{\xi_m(t) + p_m^{\mathrm{av}}}{n_m}\right) + \sum_{i=1}^{M}\beta d_i(t). \qquad (14)$$

An optimal solution to (5) and also (14) has $d_m(t) = 0$ for all m. That is, the objective (14) can equivalently be replaced by the objective of minimizing $-\sum_{m=1}^{M}\log\left(1 + \dfrac{\xi_m(t) + p_m^{\mathrm{av}}}{n_m}\right)$ and by adding the constraint

$d_m(t) = 0$ for all m. This is because the penalty for dropping is β. Thus, it can be shown that it is always better to restrict data at the Upper Network Layer (or even at transport layer) rather than admitting it and later dropping it. We recommend choosing β such that $1 \leq \beta \leq 2$. Because larger value of β will trade packet drops at the Lower Network Layer for packet non-admissions at the flow controller. Furthermore, assuming that: $u_{0\min}{}^m$ is deterministic low bound on $u_0{}^m(t)$ for all m in each slot t, the $g_{\min}{}^0$ is a deterministic lower bound on $g(\mathbf{p}(t),\mathbf{d}(t), \alpha_{NL}(t), t)$ for all t, plus assuming that there is a finite constant $K > 0$ such that for all choices of $\alpha_{NL}(t)$ on slot t we have:

$$\mathbf{E}\left\{ A_m(\alpha_{NL}(t),\psi(t))^4 \right\} \leq K \ \ \forall m \in \{1,2,..,M\}, \quad (15)$$

$$\mathbf{E}\left\{ \mu_m(\alpha_{NL}(t),\psi(t))^4 \right\} \leq K \ \ \forall m \in \{1,2,..,M\}, \quad (16)$$

$$\mathbf{E}\left\{ u_m(\alpha_{NL}(t),\psi(t))^4 \right\} \leq K \ \ \forall m \in \{1,2,..,M\}, \quad (17)$$

$$\mathbf{E}\left\{ d_m(\alpha_{NL}(t),\psi(t))^4 \right\} \leq K \ \ \forall m \in \{1,2,..,M\}, \quad (18)$$

$$\mathbf{E}\left\{ \xi_m(\alpha_{NL}(t),\psi(t))^4 \right\} \leq K \ \ \forall m \in \{1,2,..,M\} \quad (19)$$

where the expectations are taken with respect to the distribution of the i.i.d. $\psi(t)$ process, and possibility randomized decisions $\alpha_{NL}(t) \in A_{\psi(t)}$. The expressions (15)–(19) are implying all major forms of stability and further are helping in the process of determining the algorithm's performance bounds. By squaring (2), (12) and (15); then summarizing according to the (16), doing the one-step conditional Lyapunov drift (given in Sec. 4.5 in [10]) and then adding the "penalties" $V_1 \cdot \mathbf{E}\{\mathbf{u}(\mathbf{x}(t), \alpha_{NL}(t), t)|\mathbf{S}(t)\}$ and $V_2 \cdot \mathbf{E}\{g(\mathbf{p}(t),\mathbf{d}(t), \alpha_{NL}(t), t)|\mathbf{S}(t)\}$ to both sides, it is not difficult to show that the expression (13) has the following upper bound for all t (for more details see the Theorem 4.2 in [10]):

$$B - V_1 \mathbf{E}\left\{ \sum_{m=1}^{M}\sum_{i=1}^{N} w_{i,m} x_i(t) \Big| \mathbf{S}(t) \right\} - $$

$$V_2 \mathbf{E}\left\{ \sum_{m=1}^{M}\left(\log\left(1 + \frac{\xi_m(t) + p_m{}^{av}}{n_m}\right) - \beta d_m(t) \right) \Big| \mathbf{S}(t) \right\} + $$

$$\sum_{m=1}^{M} Q_m(t)\mathbf{E}\left\{ \sum_{i=1}^{N} w_{i,m} x_i(t) - \mu_m(t) - d_m(t) \Big| \mathbf{S}(t) \right\} + \quad (20)$$

$$\sum_{m=1}^{M} U_m(t)\mathbf{E}\left\{ \mu_{\min} - \mu_m(t) - d_m(t) \Big| \mathbf{S}(t) \right\} + $$

$$\sum_{l=1}^{M} Z_l(t)\mathbf{E}\left\{ \xi_l(t) \Big| \mathbf{S}(t) \right\}.$$

Here B is a finite constant related to the worst-case second moments of $A_m(t)$, $\xi_m(t)$, $d_m(t)$ and $\mu_m(t)$ processes:

$$B \triangleq \frac{1}{2}\left(\sum_{m=1}^{M} \mathbf{E}\left\{ \sum_{i=1}^{N}\left((w_{i,m} x_i(t))^2 + \mu_m{}^2(t) + d_m{}^2(t) \right) \Big| \mathbf{S}(t) \right\} \right) + $$

$$\frac{1}{2}\left(\sum_{m=1}^{M} \mathbf{E}\left\{ \left(\mu_{\min}{}^2 + \mu_m{}^2(t) + d_m{}^2(t) \right) \Big| \mathbf{S}(t) \right\} \right) + \quad (21)$$

$$\frac{1}{2}\left(\sum_{l=1}^{M} \mathbf{E}\left\{ \xi_m{}^2(t) \Big| \mathbf{S}(t) \right\} \right).$$

Such a constant B exists by the boundedness assumptions on the processes. Furthermore, the goal of our MassiveAQUA algorithm is to minimize the expression (20), by choosing the most appropriate control policy action $\alpha_{NL}(t) \in A_{\psi(t)}$ every slot t, observing the real and virtual queues vectors $\mathbf{Q}(t)$, $\mathbf{Z}(t)$, $\mathbf{U}(t)$ and the current state $\psi(t)$. In that way, the MassiveAQUA algorithm is solving the above discussed optimization problem, decoupling and reducing it into separate algorithms and does the following:

1) Stream control and QoS traffic routing over different RATs interfaces at the 5GTs (vertical multi-homing and multi-streaming [8],[9]): Every slot t, for each queue $m \in \{1,2,3,\dots, M\}$ observes new arrivals $x_m(t)$, the actual queue $Q_m(t)$ (2) and the virtual queues $Z_m(t)$ (12) and $U_m(t)$ (15), and chooses $x_m(t)$ to go over that queues to:

$$\text{Minimize:} \sum_{m=1}^{M} Q_m(t)\sum_{i=1}^{N} w_{i,m} x_i(t) - V_1 \cdot \sum_{j=1}^{M}\sum_{i=1}^{N} w_{i,j} x_i(t)$$

$$\text{subject to: } 0 \leq \sum_{i=1}^{N} w_{i,m} x_i(t) \leq \mu_m{}^{\max}, \forall m \in \{1,2,...,M\} \quad (22)$$

$$\text{and} \quad \sum_{j=1}^{M} w_{i,j} = 1, \ \forall i \in \{1,2,...,N\}.$$

2) Average Power Allocation: Every slot t, observe virtual queue $Z_m(t)$ for $m \in \{1,2,3,\dots, M\}$ and allocate power vector $\mathbf{p}(t)$ on each interface (so there is power-aware uplink/downlink transmission scheduling), as the solution for the auxiliary variable $\xi_m(t)$ to:

$$\text{Minimize:} \sum_{l=1}^{M} Z_l(t)\xi_l(t) - V_2 \sum_{m=1}^{M} \log\left(1 + \frac{p_m{}^{av} + \xi_m(t)}{n_m}\right) \quad (23)$$

$$\text{subject to: } -p_m{}^{av} \leq \xi_m(t) \leq \xi_m{}^{\max}, \forall m \in \{1,2,...,M\}.$$

3) Packet Dropping: For every slot t, observe virtual queue $U_m(t)$ (11) and real queue $Q_m(t)$ (1), and for $m \in \{1,2,3,\dots, M\}$ choose $d_m(t)$ by:

$$d_m(t) = \begin{cases} A_{\max}, & \text{if } U_m(t) + Q_m(t) > \beta V_2 \\ 0, & \text{else} \end{cases} \quad (24)$$

4) Queues Updates: For $m \in \{1,2,3,\dots, M\}$ update actual queues $Q_m(t)$, virtual queues $Z_m(t)$ and $U_m(t)$ according to (2), (12) and (15), respectively.

5) Downlink packet routing: IP packets, which are coming from all RATs interfaces and going to the application layer, are received in the incoming recipient's (another additional actual and virtual) queues, and sent from MassiveAQUA module to the Upper Network Layer. Then all packets are resent to Transport Layer where they are synthesized in one stream (flow) $x_m(t)$ and finally delivered to the peer application. As before mentioned, the arrivals, $\psi(t)$ process and channel states are assumed to be i.i.d. every slot t, and the queue network is assumed to be initially empty at time 0.

5. Performance Bounds

Our novel 5GMT is multi-RAT node, with several (n) interfaces, each for different RAT (overall n RATs). Assuming there is a finite constant $C \geq 0$ such that our MassiveAQUA algorithm chooses the most appropriate control policy action $\alpha_{NL}(t) \in A_{\psi(t)}$ every slot t, observing the real and virtual queues vectors $\mathbf{Q}(t)$, $\mathbf{Z}(t)$, $\mathbf{U}(t)$ i.e. with combining queue vector $\mathbf{S}(t)$, and minimizing the upper bound (25), then we have the following theorem:

Theorem 1 (Performance Bounds of Massive-AQUA): If we suppose that the arrivals and channel state information for each time slots is stochastic i.i.d. over slots with probabilities $\pi(\psi)$, the problem for maximizing $\sum_{m=1}^{M}\sum_{i=1}^{N}w_{i,m}\overline{x_i(t)}$ and (7-9) are feasible, and if we fixed a value for the constant $C \geq 0$, using C-additive approximation of the MassiveAQUA algorithm every slot t, and having appropriate control policy action ($\alpha_{NL}(t)$), plus setting the penalty control parameters $V_1 = V_2 = V$ then:

1) The time average expected cost satisfies:

$$\limsup_{t \to \infty} \frac{1}{t}\sum_{\tau=0}^{t-1}\left(\sum_{m=1}^{M}\mathbf{E}\left\{u_0^{\ m}(\tau)\right\} + \mathbf{E}\left\{g(\tau)\right\}\right) \tag{25}$$
$$\geq X_{\Sigma}^{\text{optim}} - \frac{B+C}{V} - g^{\text{optim}}$$

where $X_{\Sigma}^{\text{optim}}$ is the optimal infimum time average of the weighted sum of admitted rates into the queue network, achievable by any policy that stabilizes the queues and satisfied average power constraints and all other required constraints. The g^{optim} is the optimal infimum time average of the objective function (19), also achievable by any policy that stabilizes the queues and satisfied average power constraints and all other required constraints. Also B is defined with (21).

2) If we assume that there is a finite constraints $K > 0$, $\rho \geq 0$ and $\delta > 0$ for all (possibly random) choices of the control policy action $\alpha_{NL}(t)$ where (15)–(19) are met and the following constraints are satisfied for all m:

$$\mathbf{E}\left\{\xi_m(\alpha_{NL}(t),t)\right\} \leq -\delta + \rho , \tag{26}$$

$$\mathbf{E}\left\{d_m(\alpha_{NL}(t),t)\right\} \leq -\delta + \rho , \tag{27}$$

$$\mathbf{E}\left\{A_m(\alpha_{NL}(t),t) - \mu_m(\alpha_{NL}(t),t) - d_m(\alpha_{NL}(t),t)\right\} \leq -\delta + \rho, \tag{28}$$

$$\mathbf{E}\left\{\sum_{m=1}^{M}\sum_{i=1}^{N}w_{i,m}x_i(\alpha_{NL}(t),t)\right\} \leq X_{\Sigma}^{\text{optim}} + \rho. \tag{29}$$

$$\mathbf{E}\left\{\sum_{m=1}^{M}g(\alpha_{NL}(t),t)\right\} \leq g^{\text{optim}} + \rho. \tag{30}$$

Then: $\tag{31}$

$$\limsup_{t \to \infty}\frac{1}{t}\sum_{\tau=0}^{t-1}\sum_{m=1}^{M}\mathbf{E}\left\{Q_m(\tau) + Z_m(\tau) + U_m(\tau)\right\} \leq \frac{B+C}{\delta} -$$
$$- \frac{V(X_{\Sigma}^{\text{optim}} + u^{\min} - g^{\text{optim}} + g_{\min}^{\ 0})}{\delta}$$

where

$$u^{\min} \leq \frac{1}{t}\sum_{\tau=0}^{t-1}\sum_{m=1}^{M}\mathbf{E}\left\{u_0^{\ m}(\alpha_{NL}(\tau),\tau)\right\} \leq u^{\max}, \forall t \in \{1,2,...,t-1\} \text{ , and}$$

$$g_{\min}^{\ 0} \leq \frac{1}{t}\sum_{\tau=0}^{t-1}\mathbf{E}\left\{g(\alpha_{NL}(\tau),\tau)\right\} \leq g_{\max}^{\ 0}, \forall t \in \{1,2,...,t-1\} \cdot$$

3) The real and virtual queue vectors $\mathbf{Q}(t)$, $\mathbf{U}(t)$ and $\mathbf{Z}(t)$ are mean rate stable, and all required constraints of the optimization problem (5–9) are satisfied.

Proof. (Theorem 1)

Due to the choice of the most appropriate control policy action $\alpha_{NL}(t) \in A_{\psi(t)}$ every slot t, by the Massive-AQUA algorithm, observing the combining queue vector $\mathbf{S}(t)$, finite constant $C \geq 0$, which choice is minimizing (25), setting the penalty control parameters $V_1 = V_2 = V$, we have for all t and all possible vectors $\mathbf{S}(t)$:

$$\Delta(\mathbf{S}(t)) + V \cdot \mathbf{E}\left\{\mathbf{u}(\mathbf{x}(t),\alpha_{NL}(t),t)\big|\mathbf{S}(t)\right\} +$$
$$V \cdot \mathbf{E}\left\{g(\mathbf{p}(t),\mathbf{d}(t),\alpha_{NL}(t),t)\big|\mathbf{S}(t)\right\} \leq$$
$$B + C - V\mathbf{E}\left\{\sum_{m=1}^{M}\sum_{i=1}^{N}w_{i,m}x_i(\alpha_L^{\ e}(t),t)\big|\mathbf{S}(t)\right\} -$$
$$V\mathbf{E}\left\{\sum_{m=1}^{M}\left(\log\left(1 + \frac{\xi_m(\alpha_L^{\ e}(t),t) + p_m^{\ av}}{n_m}\right) - \beta d_m(\alpha_L^{\ e}(t),t)\right)\big|\mathbf{S}(t)\right\} +$$
$$\sum_{m=1}^{M}Q_m(t)\mathbf{E}\left\{\sum_{i=1}^{N}w_{i,m}x_i(\alpha_L^{\ e}(t),t) - \mu_m(\alpha_L^{\ e}(t),t) - d_m(\alpha_L^{\ e}(t),t)\big|\mathbf{S}(t)\right\} +$$
$$\sum_{m=1}^{M}U_m(t)\mathbf{E}\left\{\mu_{\min} - \mu_m(\alpha_L^{\ e}(t),t) - d_m(\alpha_L^{\ e}(t),t)\big|\mathbf{S}(t)\right\} +$$
$$\sum_{l=1}^{M}Z_l(t)\mathbf{E}\left\{\xi_l(\alpha_L^{\ e}(t),t)\big|\mathbf{S}(t)\right\}. \tag{32}$$

where $\alpha_L^{\ e}(t)$ is another decision action that can be implemented on a slot t. If we fixed δ in the interval $(0, \delta_{\max}]$, and $\rho > 0$, plus using the control policy action $\alpha_L^{\ e}(t)$ designed to achieve (26)–(30) and nothing that this policy action makes decisions independent of the $\mathbf{S}(t)$ (the values $x_i(\alpha_{NL}(t),t)$, $\mu_m(\alpha_{NL}(t),t)$, $d_m(\alpha_{NL}(t),t)$, and $\xi_m(\alpha_{NL}(t),t)$ are independent of current queue backlog and state $\mathbf{S}(t)$), yields:

$$\Delta(\mathbf{S}(t)) + V \cdot \mathbf{E}\left\{\mathbf{u}(\mathbf{x}(t),\alpha_{NL}(t),t)\big|\mathbf{S}(t)\right\} +$$
$$V \cdot \mathbf{E}\left\{g(\mathbf{p}(t),\mathbf{q}(t),\alpha_{NL}(t),t)\big|\mathbf{S}(t)\right\} \leq$$
$$B + C - V\left(X_{\Sigma}^{\text{optim}} + \rho\right) + V\left(g^{\text{optim}} + \rho\right) +$$
$$(\rho - \delta)\sum_{m=1}^{M}U_m(t) + (\rho - \delta)\sum_{m=1}^{M}Q_m(t) + (\rho - \delta)\sum_{l=1}^{M}Z_l(t). \tag{33}$$

Also, the (33) holds for all $\rho > 0$. Now if we take the limit as $\rho \to 0$ yields:

$$\Delta(\mathbf{S}(t)) + V \cdot \mathbf{E}\left\{\mathbf{u}(\mathbf{x}(t),\alpha_{NL}(t),t)\big|\mathbf{S}(t)\right\} +$$
$$V \cdot \mathbf{E}\left\{g(\mathbf{p}(t),\mathbf{q}(t),\alpha_{NL}(t),t)\big|\mathbf{S}(t)\right\} \leq \tag{34}$$
$$B + C - V\left(X_{\Sigma}^{\text{optim}}\right) + V\left(g^{\text{optim}}\right) -$$
$$\delta\sum_{m=1}^{M}\left(Q_m(t) + U_m(t) + Z_m(t)\right).$$

Furthermore, let the fourth moment bounds (15)–(19) are holding and applying Theorem 4.2 (Lyapunov Optimization) from [10], so the constraints (6)–(9) are satisfied, by using the low of telescoping sums over $\tau \in \{0,1,2,\ldots,t-1\}$ for $t > 0$, the (34) yields:

$$
\mathbf{E}\{L(\mathbf{S}(t))\} - \mathbf{E}\{L(\mathbf{S}(0))\} +
$$
$$
V \cdot \sum_{\tau=0}^{t-1} \sum_{m=1}^{M} \mathbf{E}\{u_0{}^m(\tau)\} + V \cdot \sum_{\tau=1}^{t-1} \mathbf{E}\{g(\tau)\} \le \tag{35}
$$
$$
\left(B + C - V \cdot X_\Sigma^{\text{optim}} + V \cdot g^{\text{optim}}\right) \cdot t -
$$
$$
\delta \sum_{\tau=0}^{t-1} \sum_{m=1}^{M} (Q_m(\tau) + U_m(\tau) + Z_m(\tau)).
$$

Moreover, rearranging terms and neglecting non-negative terms when appropriate, it is easy to show that the inequality (35) directly implies the following two inequalities for all $t > 0$:

$$
\frac{1}{t} \cdot \sum_{\tau=0}^{t-1} \left(\sum_{m=1}^{M} \mathbf{E}\{u_0{}^m(\tau)\} + \mathbf{E}\{g(\tau)\} \right) \le \left(\frac{B+C}{V} - X_\Sigma^{\text{optim}} + g^{\text{optim}} \right) -
$$
$$
\frac{\delta}{V \cdot t} \sum_{\tau=0}^{t-1} \sum_{m=1}^{M} (Q_m(\tau) + U_m(\tau) + Z_m(\tau)) + \frac{\mathbf{E}\{L(\mathbf{S}(0))\} - \mathbf{E}\{L(\mathbf{S}(t))\}}{V \cdot t}. \tag{36}
$$

and

$$
\frac{1}{t} \sum_{\tau=0}^{t-1} \sum_{m=1}^{M} (Q_m(\tau) + U_m(\tau) + Z_m(\tau)) \le \frac{B+C}{\delta} -
$$
$$
\frac{V}{\delta} \left(X_\Sigma^{\text{optim}} - g^{\text{optim}} + \frac{1}{t} \cdot \sum_{\tau=0}^{t-1} \sum_{m=1}^{M} \mathbf{E}\{u_0{}^m(\tau)\} + \frac{1}{t} \cdot \sum_{\tau=0}^{t-1} \mathbf{E}\{g(\tau)\} \right) + \tag{37}
$$
$$
\frac{\mathbf{E}\{L(\mathbf{S}(0))\} - \mathbf{E}\{L(\mathbf{S}(t))\}}{\delta \cdot t}.
$$

where (36) follows by dividing (35) by V and t, and (37) follows by dividing (35) by δ and t. Now, recalling that the time average expectation for the penalty sum of all $u_0{}^m(\alpha_{\text{NL}}(t),t)$ is deterministically lower bounded by $u^{\min}$ (possibly negative value), also the time average expectation for the objective function (14) is deterministically lower bounded by $g_{\min}{}^0$ (possible negative value) and finally, taking limits of the (42) as $t \to \infty$ proves (31). Moreover, by taking limits of the (36) as $t \to \infty$ plus multiplying the result inequality by -1 proves (25).

6. Simulation Results and Analysis

In this section, we provide a simulation results and analysis of the presented MassiveAQUA algorithm for a scenario including total wireless and mobile network coverage from several RATs, for the 250 5G mobile terminals and one 5G fixed terminal (server at the core network). From the three different scenarios, we are obtaining the simulation results for average throughput, as well as the average queue backlog values for different network conditions, together with the average queue delays of the 5GTs. Our simulation scenarios are multi-cell case, plotted in the dense urban area, with random initial locations of mobile terminals uniformly distributed within the entire urban area. To emphasize that the used maximal output rates are carefully chosen to be adequate to the maximal uplink serving rates (by dividing the uplink bit rates with the average number of bits per packet) of LTE, LTE-Advanced, IEEE 802.11n, IEEE 802.16e, IEEE 802.16m, IEEE 802.11ac and IEEE 802.11b RATs, shared between 250 mobile terminals in one cell. In Tab. 1 the most important

Scenario	Parameters and values	Description
Scenario 1	$v = 40$ km/h	Average velocity of the mobile terminals
	$\mathbf{x}_{10}$=[68.9655; 50; 24.316; 100; 50; 50; 24.316; 50; 68.9655; 68.9655] $\mathbf{x}_9$=[68.9655; 50; 24.316; 100; 50; 50; 24.316; 50; 68.9655] $\mathbf{x}_8$=[68.9655; 50; 24.316; 100; 50; 50; 24.316; 50] $\mathbf{x}_7$=[68.9655; 50; 24.316; 100; 50; 50; 24.316] $\mathbf{x}_6$=[68.9655; 50; 24.316; 100; 50; 50] $\mathbf{x}_5$=[68.9655; 50; 24.316; 100; 50] $\mathbf{x}_4$=[68.9655; 50; 24.316; 100] $\mathbf{x}_3$=[68.9655; 50; 24.316] $\mathbf{x}_2$=[68.9655; 50] $\mathbf{x}_1$=[68.9655]	Vectors of the time average arrival rates for different number of services
	μ_1= (1/250) [44643]; μ_2= (1/250) [44643; 12500]; μ_3= (1/250) [6694.4; 37000; 12500]; μ_4= (1/250) [6694.4; 12500; 5000; 44643]; μ_5= (1/250) [6694.4; 12500; 5000; 44643; 5000]; μ_6= (1/250) [6694.4; 12500; 5000; 44643; 12500; 6694.4].	Vectors of the output serving rates for different number of RAT interfaces (M queues).
Scenario 2	μ_1= (1/250) [6694.4]; μ_2= (1/250) [6694.4; 12500]; μ_3= (1/250) [6694.4; 5000; 12500]; μ_4= (1/250) [6694.4; 12500; 5000; 6694.4]; μ_5= (1/250) [6694.4; 12500; 5000; 12053.57; 2250].	Vector of the output serving rates for different number (M) of RAT interfaces.
	$\mathbf{x}$=[68.9655; 50; 24.316]	Vector for the time average arrival rates for N=3 services
Scenario 1 & 2	$V = V_1 = V_2 = 10$	Penalty control parameters
Scenario 3	$\mathbf{x}$=[69; 50]	Vector for the time average arrival rates for N=2 services
	$v = 40$ km/h	Average velocity of the mobile terminals
	For Rout=76.7776: μ_2= (1/250) [6694.4; 12500]; For Rout=93.4456: μ_2= (1/250) [6694.4; 16667]; For Rout=96.7776: μ_3= (1/250) [6694.4; 12500 5000]; For Rout=98.2160: μ_2= (1/250) [12054; 12500]; For Rout=103.5552: μ_3= (1/250) [6694.4; 12500; 6694.4].	Vectors of the output serving rates for different number of RAT interfaces (M queues) and different aggregated throughput.
	V_1=10	First penalty control parameter
Scenario 1,2,3	$p_m{}^{\text{av}}$ =0.7 W	Average power for each interface (W)

Tab. 1. Simulation parameters for scenario 1, 2 and 3.

simulation parameters for all three scenarios are presented. Each simulation was run over 10^7 time slots. In Fig. 2 the average queue backlog versus different number of multimedia sources is presented. The time average arrival rates for different number of used services (applications) are given in Tab. 1. Undoubtedly, as the number of used interfaces in 5GTs is higher, the average queue backlog is becoming smaller. The case when they are 6 interfaces is achieving superior results for any number of data sources over other four cases (especially with the case when there is only one used interface (for $M = 1$ in Fig. 2)). By the Little's theorem [33], dividing the average queuing backlog with the average arrival rates yields an upper bound on average time delay. Consequently, in the case with more interfaces within 5GT with MassiveAQUA algorithm will cause minimal average time delay for any used service. Furthermore, Fig. 3 presents the average uplink throughput per user versus the average velocity of the 5G mobile terminals, when the number of sources is fixed ($N = 3$). This is Scenario 2 and the services are: VoIP-conference, VoIP and delay-sensitive data service. It is evident that as the number of RAT interfaces is higher (i.e. for $M = 5$) the throughput is higher for each velocity, even it is with diminishing trend versus velocity. Also, with smaller number of interfaces the average throughput is smaller and is supporting only those RAT interfaces belonging to the RAT with the best mobility support. A key observation about the above algorithm is that it does optimal flow control decisions and simultaneously causing network queuing stability and minimal network queue backlog. As expected, the worst case uplink throughput we have when there are mobile terminals with one interface ($M = 1$). In Scenario 3, we have only two sources of information (two services), and the capacity region is shown in Fig. 4, together with the illustration for the parameter ro (For $\lambda = (50, 69)$ (i.e., point D illustrated) different lengths for the segment OA are allowed (the parameter ro is OA/OD)). Moreover, Fig. 5 shows the average queue delays versus different values of the parameter ro, for different values of the penalty parameter V_2, when the penalty parameter for the first penalty function is $V_1 = 10$. As the parameter ro is increasing, it is near it's maximal value (which is 1, i.e. when the segment OA = OD in Fig. 4), so the average delays are becoming more stable and bounded. As can be notice from Fig. 5, when we set the penalty control parameter half of the previous value (i.e. when $V_2 = 5$), in comparison with the case when the V_2 is 10, the delays for all five cases are becoming lower with maximal values around 0.15 slots. At the same time, it tends to push the queues towards a lower congestion state, but we get large packet dropping ratio (which is proportional with the decreasing of the control parameter V_2, according to (24)). Thus, we minimize a weighted sum of drift and "penalty" for the second (novel) penalty function (14), decreasing the delay, but we are increasing the packet dropping ratio for all queues. For both values of penalty parameter V_2 (i.e. when $V_2 = 10$ and when $V_2 = 5$), the case when we have Rout = 98.2160, i.e.
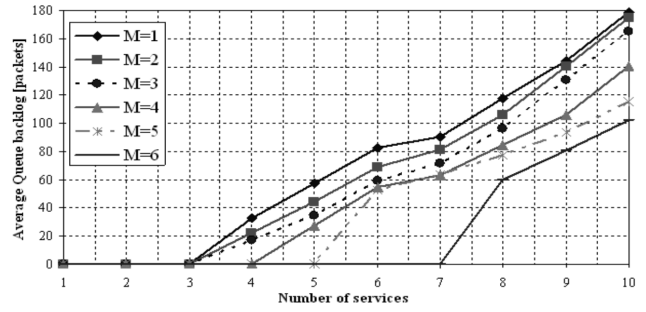

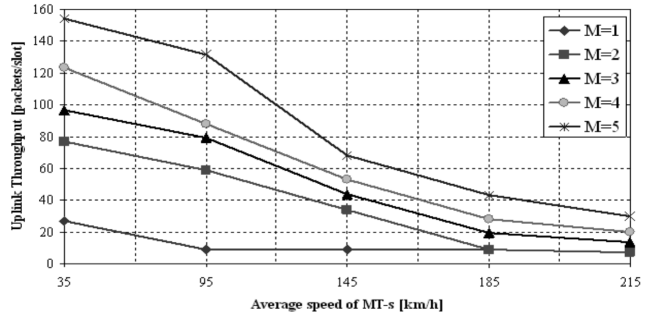**Fig. 2.** The average queue backlog versus number of sources.

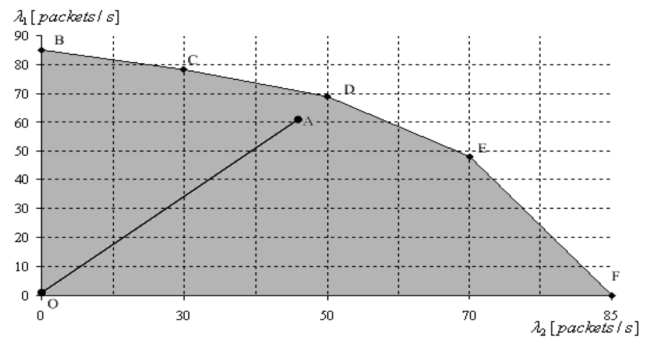**Fig. 3.** The average throughput vs. average velocity of MTs.

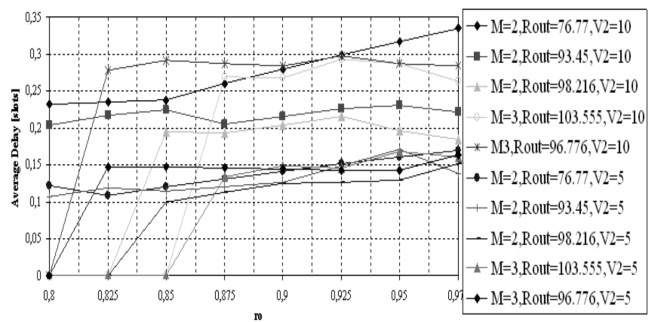**Fig. 4.** The capacity region Λ.

Fig. 5. The average delay versus parameter ro.

balanced queue with almost equal values for their output serving rates (i.e. when $\mu_2 = (1/250)$ [12054; 12500]), we have the lowest delay in comparison with all other cases, even when we have three RAT interfaces within 5GTs. Thus, we can conclude that the values for the output serving rates of the RAT interfaces are having very important impact in the performances of the MassiveAQUA algorithm, and should be quite balanced.

7. Conclusion

This paper presents joint stochastic Lyapunov stability and performance optimization, and develops utility optimal traffic control for the future 5G mobile terminals and 5G fixed terminals (proxy servers in the mobile core network part). It defines a novel analytical framework and performance bounds for 5GT with vertical multi-homing and multi-streaming features for 5G broadband networks, which maximizes the time average throughput and at the same time stabilizes the queuing, simultaneously providing minimal queue backlog and delay for each real-time delay-sensitive application. The 5GTs with Lyapunov drift-plus-penalty technique is implemented in user MTs and in the core network (as 5G proxy server), and it is handling simultaneously multiple multimedia real-time applications via multiple wireless/mobile network interfaces with optimal packet scheduling and high level of QoS provisioning.

References

[1] JANEVSKI, T. *Internet Technologies for Fixed and Mobile Networks*. USA: Artech House, 2015, ISBN: 978-1-60807-921-6

[2] BOCCARDI, F., et al. Five disruptive technology directions for 5G. *IEEE Communications Magazine*, 2014, vol. 52, no. 2, p. 74 to 80. DOI: 10.1109/MCOM.2014.6736746

[3] BANGERTER, B., et al. Networks and devices for the 5G era. *IEEE Communications Magazine*, 2014, vol. 52, no. 2, p. 90–96. DOI: 10.1109/MCOM.2014.6736748

[4] CHENG-XIANG WANG, et al. Cellular architecture and key technologies for 5G wireless communication networks. *IEEE Communications Magazine*. 2014, vol. 52, no. 2, p. 122–130. DOI: 10.1109/MCOM.2014.6736752

[5] JANEVSKI, T. 5G Mobile phone concept. In *IEEE Consumer Communications and Networking Conference (CCNC)*. Las Vegas (USA), 2009. DOI: 10.1109/CCNC.2009.4784727

[6] WANG, P., et al. Multi-Gigabit millimeter wave wireless communications for 5G: From fixed access to cellular networks. *IEEE Communications Magazine*, 2015, vol. 53, no. 1, p. 168–178. DOI: 10.1109/MCOM.2015.7010531

[7] QIAO, J., et al. Enabling device-to-device communications in millimeter-wave 5G cellular networks. *IEEE Communications Magazine*, 2015, vol. 53, no. 1, p. 209–215. DOI: 10.1109/MCOM.2015.7010536

[8] Rec. ITU-T *Y.2052* (02/2008): Framework of multi-homing in IPv6-based NGN, February 2008.

[9] Rec. ITU-T *Y.2056* (08/2011): Framework of vertical multihoming in IPv6-based Next Generation Networks, August 2011.

[10] NEELY, M. J. *Stochastic Network Optimization with Application to Communication and Queueing Systems*. USA: Morgan & Claypool, 2010. DOI: 10.2200/S00271ED1V01Y201006CNT007

[11] MALISOFF, M., MAZENC, F. *Constructions of Strict Lyapunov Functions*. London: Springer, 2009. ISBN: 978-1-84882-535-2

[12] TASSIULAS, L., EPHREMIDES, A. Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE Transactions on Automatic Control*, Dec. 1992, vol. 37, no. 12, p. 1936–1948. DOI: 10.1109/9.182479

[13] TASSIULAS, L., EPHREMIDES, A. Dynamic server allocation to parallel queues with randomly varying connectivity. *IEEE Transactions on Information Theory*, March 1993, vol. 39, no. 2, p. 466–478. DOI: 10.1109/18.212277

[14] KUMAR, P. R., MEYN, S. P. Stability of queueing networks and scheduling policies. *IEEE Transactions on Automatic Control*, Feb. 1995, vol.40, no. 2, p. 251–260. DOI: 10.1109/9.341782

[15] NEELY, M. J., MODIANO, E., LI, C. Fairness and optimal stochastic control for heterogeneous networks. In *Proceedings of IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies INFOCOM*. Miami (FL, USA), March 2005. DOI: 10.1109/INFCOM.2005.1498453

[16] STOLYAR, A. Maximizing queueing network utility subject to stability: Greedy primal-dual algorithm. *Queueing Systems*, 2005, vol. 50, no. 4, p. 401–457. DOI: 10.1007/s11134-005-1450-0

[17] NEELY, M. J., MODIANO, E., ROHRS, C. E. Dynamic power allocation and routing for time varying wireless networks. *IEEE Journal on Selected Areas in Communications*, Jan. 2005, vol. 23, no. 1, p. 89–103. DOI: 10.1109/JSAC.2004.837349

[18] SRIDHARAN, A., MOELLER, S., KRISHNAMACHARI, B. Making distributed rate control using Lyapunov drifts a reality in wireless sensor networks. In *6th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks and Workshops WiOPT 2008*. Berlin (Germany), 2008, p. 452–461. DOI: 10.1109/WIOPT.2008.4586106

[19] GEORGIADIS, L., NEELY, M. J., TASSIULAS, L. Resource allocation and cross-layer control in wireless networks. *Foundations and Trends in Networking*, 2006, vol. 1, no. 1, p. 1 to 149. DOI: 10.1561/1300000001

[20] NEELY, M. J. Optimal backpressure routing for wireless networks with multi-receiver diversity. In *40th Annual Conference on Information Sciences and Systems (CISS)*. Princeton (NJ, USA), 2006, p. 18–25. DOI: 10.1109/CISS.2006.286424

[21] LEE, J. W., MAZUMDAR, R. R., SHROFF, N. B. Opportunistic power scheduling for dynamic multiserver wireless systems. *IEEE Transactions on Wireless Communications*, June 2006, vol. 5, no. 6, p. 1506–1515. DOI: 10.1109/TWC.2006.1638671

[22] NEELY, M. J. Energy optimal control for time varying wireless networks. *IEEE Transactions on Information Theory*, July 2006, vol. 52, no. 7, p. 2915–2934. DOI: 10.1109/TIT.2006.876219

[23] NEELY, M. J. Super-fast delay tradeoffs for utility optimal fair scheduling in wireless networks. *IEEE Journal on Selected Areas in Communications*, Aug. 2006, vol. 24, no. 8, p. 1489–1501. DOI: 10.1109/JSAC.2006.879357

[24] ERYILMAZ, A., SRIKANT, R. Joint congestion control, routing, and mac for stability and fairness in wireless networks. *IEEE Journal on Selected Areas in Communications*, Special Issue on Nonlinear Optimization of Communication Systems, Aug. 2006, vol. 14, p. 1514–1524. DOI: 10.1109/JSAC.2006.879361

[25] NEELY, M. J., URGAONKAR, R. Cross-layer adaptive control for wireless mesh networks. *Ad Hoc Networks (Elsevier journal)*, August 2007, vol. 5, no. 6, p. 719–743. Available at: http://dx.doi.org/10.1016/j.adhoc.2007.01.004

[26] NEELY, M. J. Optimal energy and delay tradeoffs for multi-user wireless downlinks. *IEEE Transactions on Information Theory*, Sept. 2007, vol. 53, no. 9, p. 3095–3113. DOI: 10.1109/TIT.2007.903141

[27] ERYILMAZ, A., SRIKANT, R. Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control. *IEEE/ACM Transactions on Networking*, 2007, vol. 15, no. 6, p. 1333–1344. DOI: 10.1109/TNET.2007.897944.

[28] LI, Q., NEGI, R. Scheduling in wireless networks under uncertainties: A greedy primal-dual approach. *Arxiv Technical Report*: arXiv:1001:2050v2, June 2010.

[29] NEELY, M. J., URGAONKAR, R. Optimal backpressure routing in wireless networks with multi-receiver diversity. *Ad Hoc Networks (Elsevier journal)*, July 2009, vol. 7, no. 5, p. 862–881. DOI: 10.1016/j.adhoc.2008.07.009

[30] SHUMINOSKI, T., JANEVSKI, T. Lyapunov optimization framework for 5G mobile nodes with multi-homing. *IEEE Communications Letters*, May 2016, vol. 20, no. 5, p. 1026–1029. DOI: 10.1109/LCOMM.2016.2540622

[31] SHUMINOSKI, T., JANEVSKI, T. 5G mobile terminals with advanced QoS-based user-centric aggregation (AQUA) for heterogeneous wireless and mobile networks. *Wireless Networks*, 2016, vol. 22, no. 5, p. 1553–1570. DOI: 10.1007/s11276-015-1047-4

[32] SHUMINOSKI, T., JANEVSKI, T. Radio network aggregation for 5G mobile terminals in heterogeneous wireless and mobile networks. *Wireless Personal Communication*. May 2014, vol. 78, no. 2, p. 1211–1229. DOI: 10.1007/s11277-014-1813-0

[33] BERTSEKAS, D. P., GALLAGER, R. *Data Networks*. New Jersey: Prentice-Hall, 1992. ISBN 0132009161

About the Authors …

Tomislav SHUMINOSKI (corresponding author) received the B.Sc. (2008), M.Sc. (2010) and Ph.D. (2016) degrees from the Faculty of Electrical Engineering and Information Technologies, University Ss Cyril and Methodius, Skopje, Macedonia, where he is an Assistant Professor.

Toni JANEVSKI is a Full Professor at the Faculty of Electrical Engineering and Information Technologies (FEEIT), Ss. Cyril and Methodius University, Skopje, Macedonia, from which he received his Dipl. Ing., M.Sc. and Ph.D. degrees in 1996, 1999 and 2001, respectively.