Hindawi Publishing Corporation
Journal of Electrical and Computer Engineering
Volume 2016, Article ID 4067186, 10 pages
http://dx.doi.org/10.1155/2016/4067186

Research Article

A Game for Energy-Aware Allocation of Virtualized Network Functions

Roberto Bruschi,[1] Alessandro Carrega,[1,2] and Franco Davoli[1,2]

[1] *CNIT-University of Genoa Research Unit, 16145 Genoa, Italy*
[2] *DITEN, University of Genoa, 16145 Genoa, Italy*

Correspondence should be addressed to Alessandro Carrega; alessandro.carrega@unige.it

Received 2 October 2015; Revised 22 December 2015; Accepted 11 January 2016

Academic Editor: Xavier Hesselbach

Network Functions Virtualization (NFV) is a network architecture concept where network functionality is virtualized and separated into multiple building blocks that may connect or be chained together to implement the required services. The main advantages consist of an increase in network flexibility and scalability. Indeed, each part of the service chain can be allocated and reallocated at runtime depending on demand. In this paper, we present and evaluate an energy-aware Game-Theory-based solution for resource allocation of Virtualized Network Functions (VNFs) within NFV environments. We consider each VNF as a player of the problem that competes for the physical network node capacity pool, seeking the minimization of individual cost functions. The physical network nodes dynamically adjust their processing capacity according to the incoming workload, by means of an Adaptive Rate (AR) strategy that aims at minimizing the product of energy consumption and processing delay. On the basis of the result of the nodes' AR strategy, the VNFs' resource sharing costs assume a polynomial form in the workflows, which admits a unique Nash Equilibrium (NE). We examine the effect of different (unconstrained and constrained) forms of the nodes' optimization problem on the equilibrium and compare the power consumption and delay achieved with energy-aware and non-energy-aware strategy profiles.

1. Introduction

In the last few years, power consumption has shown a growing and alarming trend in all industrial sectors, particularly in Information and Communication Technology (ICT). Public organizations, Internet Service Providers (ISPs), and telecom operators started reporting alarming statistics of network energy requirements and of the related carbon footprint since the first decade of the 2000s [1]. The Global e-Sustainability Initiative (GeSI) estimated a growth of ICT greenhouse gas emissions (in $GtCO_2e$, Gt of CO_2 equivalent gases) to 2.3% of global emissions (from 1.3% in 2002) in 2020, if no Green Network Technologies (GNTs) would be adopted [2]. On the other hand, the abatement potential of ICT in other industrial sectors is seven times the size of the ICT sector's own carbon footprint.

Only recently, due to the rise in energy price, the continuous growth of customer population, the increase in broadband access demand, and the expanding number of services being offered by telecoms and providers, has energy efficiency become a high-priority objective also for wired networks and service infrastructures (after having started to be addressed for datacenters and wireless networks).

The increasing network energy consumption essentially depends on new services offered, which follow Moore's law, by doubling every two years, and on the need to sustain an ever-growing population of users and user devices. In order to support new generation network infrastructures and related services, telecoms and ISPs need a larger equipment base, with sophisticated architecture able to perform more and more complex operations in a scalable way. Notwithstanding these efforts, it is well known that most networks and networking equipment are currently still provisioned for busy or rush hour load, which typically exceeds their average utilization by a wide margin. While this margin is generally reached in rare and short time periods, the overall power consumption in today's networks remains more or less constant with respect to different traffic utilization levels.

The growing trend toward implementing networking functionalities by means of software [3] on general-purpose machines and of making more aggressive use of virtualization—as represented by the paradigms of Software Defined Networking (SDN) [4] and Network Functions Virtualization (NFV) [5]—would also not be sufficient in itself to reduce power consumption, unless accompanied by "green" optimization and consolidation strategies acting as energy-aware traffic engineering policies at the network level [6]. At the same time, processing devices inside the network need to be capable of adapting their performance to the changing traffic conditions, by trading off power and Quality of Service (QoS) requirements. Among the various techniques that can be adopted to this purpose to implement Control Policies (CPs) in network processing devices, Dynamic Adaptation ones consist of adapting the processing rate (AR) or of exploiting low power consumption states in idle periods (LPI) [7].

In this paper, we introduce a Game-Theory-based solution for energy-aware allocation of Virtualized Network Functions (VNFs) within NFV environments. In more detail, in NFV networks, a collection of service chains must be allocated on physical network nodes. A service chain is a set of one or more VNFs grouped together to provide specific service functionality and can be represented by an oriented graph, where each node corresponds to a particular VNF and each edge describes the operational flow exchanged between a pair of VNFs.

A service request can be allocated on dedicated hardware or by using resources deployed by the Service Provider (SP) that processes the request through virtualized instances. Because of this, two types of service deployments are possible in an NFV network: (i) on physical nodes and (ii) on virtualized instances.

In this paper, we focus on the second type of service deployment. We refer to this paradigm as pure NFV. As already outlined above, the SP processes the service request by means of VNFs. In particular, we developed an energy-aware solution to the problem of VNFs' allocation on physical network nodes. This solution is based on the concept of Game Theory (GT). GT is used to model interactions among self-interested players and predict their choice of strategies to optimize cost or utility functions, until a Nash Equilibrium (NE) is reached, where no player can further increase its corresponding utility through individual action (see, e.g., [8] for specific applications in networking).

More specifically, we consider a bank of physical network nodes (in this paper, we also use the terms node and resource to refer to the physical network node) performing tasks on requests submitted by players' population. Hence, in this game, the role of the players is represented by VNFs that compete for the processing capacity pool, each by seeking the minimization of an individual cost function. The nodes can dynamically adjust their processing capacity according to the incoming workload (the processing power required by incoming VNF requests) by means of an AR strategy that aims at minimizing the product of energy consumption and processing delay. On the basis of the result of the nodes' AR strategy, the VNFs' resource sharing costs assume a polynomial form in the workloads, which admits a unique NE. We examine the effect of different (unconstrained and constrained) forms of the nodes' optimization problem on the equilibrium and compare the power consumption and delay achieved with energy-aware and non-energy-aware strategy profiles.

The rest of the paper is organized as follows. Section 2 discusses the related work. Section 3 briefly summarizes the power model and AR optimization that was introduced in [9]. Although the scenario in [9] is different, it is reasonable to assume that similar considerations are valid here, too. On the basis of this result, we derive a number of competitive strategies for the VFNs' allocation in Section 4. Section 5 presents various numerical evaluations, and Section 6 contains the conclusions.

2. Related Work

We now discuss the most relevant works that deal with the resource allocation of VNFs within NFV environments. We decided not only to describe solutions based on GT, but also to provide an overview of the current state of the art in this area. As introduced in the previous section, the key enabling paradigms that will considerably affect the dynamics of ICT networks are SDN and NFV, which are discussed in recent surveys [10–12]. Indeed, SPs and Network Operators (NOs) are facing increasing problems to design and implement novel network functionalities, following rapid changes that characterize the current ISPs and telecom operators (TOs) [13].

Virtualization represents an efficient and cost-effective strategy to exploit and share physical network resources. In this context, the Network Embedding Problem (NEP) has been considered in several recent works [14–19]. In particular, the Virtual Network Embedding Problem (VNEP) consists of finding the mapping between a set of requests for virtual network resources and the available underlying physical infrastructure (the so-called *substrate*), ensuring that some given performance requirements (on nodes and links) are guaranteed. Typical node requirements are computational resources (i.e., CPU) or storage space, whereas links have a limited bandwidth and introduce a delay. It has been shown that this problem is NP-hard (it includes as subproblem the multiway separator problem). For this reason, heuristic approaches have been devised [20].

The consolidation of virtual resources is considered in [18], by taking into account energy efficiency. The problem is formulated as a mixed integer linear programming (MILP) model, to understand the potential benefits that can be achieved by packing many different virtual tasks on the same physical infrastructure. The observed energy saving is up to 30% for nodes and up to 25% for link energy consumption. Reference [21] presents a solution for the resilient deployment of network functions, using OpenStack for the design and implementation of the proposed service orchestrator mechanism.

An allocation mechanism, based on auction theory, is proposed in [22]. In particular, the scheme selects the most remunerative virtual network requests, while satisfying QoS requirements and physical network constraints. The system is split into two network substrates modeling physical and virtual resources, with the final goal of finding the best mapping of virtual nodes and links onto physical ones according to the QoS requirements (i.e., bandwidth, delay, and CPU bounds).

A novel network architecture is proposed in [23], to provide efficient coordinated control of both internal network function state and network forwarding state, in order to help operators achieve the following goals: (i) offering and satisfying tight service level agreements (SLAs); (ii) accurately monitoring and manipulating network traffic; and (iii) minimizing operating expenses.

Various engineering problems, where the action of one component has some impacts on the other components, have been modeled by GT. Indeed, GT, which has been applied at the beginning in economics and related domains, is gaining much interest today as a powerful tool to analyze and design communication networks [24]. Therefore, the problems can be formulated in the GT framework, and a stable solution for the components is obtained using the concept of equilibrium [25]. In this regard, GT has been used extensively to develop understanding of stable operating points for autonomous networks. The nodes are considered as the players. Payoff functions are often defined according to achieved connection bandwidth or similar technical metrics.

To construct algorithms with provable convergence to equilibrium points, many approaches consider network models that can be mapped to specially constructed games. Among this type of games, potential games use a real-valued function that represents the entire player set to optimize some performance metric [8, 26]. We mention Altman et al. [27], who provide an extensive survey on networking games. The models and papers discussed in this reference mostly deal with noncooperative GT, the only exception being a short section focused on bargaining games. Finally, Seddiki et al. [25] presented an approach based on two-stage noncooperative games for bandwidth allocation, which aims at reducing the complexity of network management and avoiding bandwidth performance problems in a virtualized network environment. The first stage of the game is the bandwidth negotiation, where the SP requests bandwidth from multiple Infrastructure Providers (InPs). Each InP decides whether to accept or deny the request when the SP would cause link congestion. The second stage is the bandwidth provisioning game, where different SPs compete for the bandwidth capacity of a physical link managed by a single InP.

3. Modeling Power Management Techniques

Dynamic Adaptation techniques in physical network nodes reduce the energy usage, by exploiting the fact that systems do not need to run at peak performance all the time.

Rate adaptation is obtained by tuning the clock frequency and/or voltage of processors (DVFS, Dynamic Voltage and Frequency Scaling) or by throttling the CPU clock (i.e., the clock signal is disabled for some number of cycles at regular intervals). Decreasing the operating frequency and the voltage of the node, or throttling its clock, obviously allows the reduction of power consumption and of heat dissipation, at the price of slower performance.

Advantage can be taken of these adaptation capabilities to devise control techniques based on feedback on the incoming load. One of the first proposals is that examined in a seminal paper by Nedevschi et al. [28]. Among other possibilities, we consider here the simple optimization strategy developed in [9] aimed at minimizing the product of power consumption and processing delay with respect to the processing load. The application of this control technique gives rise to quadratic dependence of the power-delay product on the load, which will be exploited in our subsequent game theoretic resource allocation problem. Unlike the cited works, our solution considers as load the requests rate of services that the SP must process. However, apart from this difference, the general considerations described are valid here, too.

Specifically, the dynamic frequency-dependent part of the processor power consumption is proportional to the clock frequency ν and to the square of the supply voltage V [29]. However, if DVFS is used, there is proportionality between the frequency and the voltage raised to some power γ, $0 < \gamma \leq 1$ [30]. Moreover, the power scaling effect induced by alternating active-idle periods in the queueing system served by the physical network node can be accounted for by multiplying the dynamic power consumption by an increasing concave function of the resource utilization of the form $(f/\mu)^{1/\alpha}$, where $\alpha \geq 1$ is a parameter and f and μ are the workload (processing requests per unit time) and the task processing speed, respectively. By taking $\gamma = 1$ (which implies a cubic relationship in the operating frequency), the overall dependence of the power consumption $\mathbf{\Phi}$ on the processing speed (which is proportional to the operating frequency) can be expressed as [9]

$$\mathbf{\Phi}(\mu) = K\mu^3 \left(\frac{f}{\mu}\right)^{1/\alpha}, \tag{1}$$

where $K > 0$ is a proportionality constant. This expression will be exploited in the optimization problems to be defined and studied in the next section.

4. System Model of Coordinated Network Management Optimization and Players' Game

We consider the situation shown in Figure 1. There are S VNFs that are represented by the incoming workload rates $\lambda_1, \ldots, \lambda_S$, competing for N physical network nodes. The workload to the jth node is given by

$$f_j = \sum_{i=1}^{S} f_j^i, \tag{2}$$

where f_j^i is VNF i's contribution. The total workload vector of player i is $\mathbf{f}^i = \mathrm{col}[f_1^i, \ldots, f_N^i]$ and, obviously,

$$\sum_{j=1}^{N} f_j^i = \lambda_i. \tag{3}$$

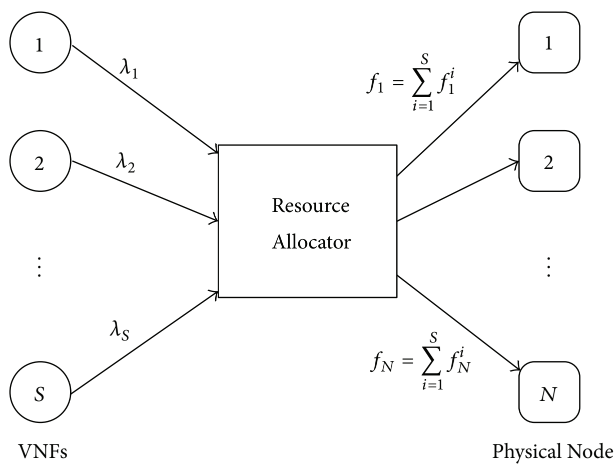

Figure 1: System model for VNFs resource allocation.

The computational resources have processing rates μ_j, $j = 1, \ldots, N$, and we associate a cost function with each one; namely,

$$c_j(\mu_j) = \frac{K_j (f_j)^{1/\alpha_j} (\mu_j)^{3-(1/\alpha_j)}}{\mu_j - f_j}. \tag{4}$$

Cost functions of the form shown in (4) have been introduced in [9] and represent the product of power and processing delay, under $M/M/1$ assumption on each node's queue. They actually correspond to the average energy consumption that can be attributed to functions' handling (considering that the node will be active whenever there are queued functions). Despite the possible inaccuracy in the representation of the queueing behavior, the denominator of (4) reflects the penalty paid for approaching the resource capacity, which is a major aspect in a model to be used for control purposes.

We now consider two specific control problems, whose resulting resource allocations are interconnected. Specifically, we assume the presence of Control Policies (CPs) acting in the network node, whose aim is to dynamically adjust the processing rate, in order to minimize cost functions of the form of (4). On the other hand, the players representing the VNFs, knowing the policy adopted by the CPs and the resulting form of their optimal costs, implement their own strategies to distribute their requests among the different resources.

The simple control structure that we envisage actually represents a situation that may be of interest in a number of different operational settings. We only mention two different cases of current high relevance: (i) a multitenant environment, where a number of ISPs are offered services by a datacenter operator (or by a telecom operator in the access network) on a number of virtual machines, and (ii) a collection of virtual environments created by a SP on behalf of its customers, where services are activated to represent customers' functionalities on their own private virtual LANs. It is worth noting that in both cases the nodes' customers may be unwilling to disclose their loads to one another, which justifies the decentralized game optimization.

4.1. Nodes' Control Policies. We consider two possible variants.

4.1.1. CP1 (Unconstrained). The direct minimization of (4) with respect to μ_j yields immediately:

$$\begin{aligned} \mu_j^*(f_j) &= \frac{3\alpha_j - 1}{2\alpha_j - 1} f_j = \vartheta_j f_j, \\ c_j^*(f_j) &= K_j \frac{(\vartheta_j)^{3-(1/\alpha_j)}}{\vartheta_j - 1} (f_j)^2 = h_j \cdot (f_j)^2. \end{aligned} \tag{5}$$

We must note that we are considering a continuous solution to the node capacity adjustment. In practice, the physical resources allow a discrete set of working frequencies, with corresponding processing capacities. This would also ensure that the processing speed does not decrease below a lower threshold, avoiding excessive delay in the case of low load. The unconstrained problem is an idealized variant that would make sense only when the load on the node is not too small.

4.1.2. CP2 (Individual Constraints). Each node has a maximum and a minimum processing capacity, which are taken explicitly into account ($\mu_j^{\min} \le \mu_j \le \mu_j^{\max}$, $j = 1, \ldots, N$). Then,

$$\mu_j^*(f_j) = \begin{cases} \vartheta_j f_j, & \underline{f}_j = \dfrac{\mu_j^{\min}}{\vartheta_j} \le f_j \le \dfrac{\mu_j^{\max}}{\vartheta_j} = \overline{f}_j, \\ \mu_j^{\min}, & 0 < f_j < \underline{f}_j, \\ \mu_j^{\max}, & \mu_j^{\max} > f_j > \overline{f}_j. \end{cases} \tag{6}$$

Therefore,

$$c_j^*(f_j) = \begin{cases} h_j \cdot (f_j)^2, & \underline{f}_j \le f_j \le \overline{f}_j, \\ K_j \dfrac{(f_j)^{1/\alpha_j} (\mu_j^{\max})^{3-(1/\alpha_j)}}{\mu_j^{\max} - f_j}, & \mu_j^{\max} > f_j > \overline{f}_j, \\ K_j \dfrac{(f_j)^{1/\alpha_j} (\mu_j^{\min})^{3-(1/\alpha_j)}}{\mu_j^{\min} - f_j}, & 0 < f_j < \underline{f}_j. \end{cases} \tag{7}$$

In this case, we assume explicitly that $\sum_{i=1}^{S} \lambda_i < \sum_{j=1}^{N} \mu_j^{\max}$.

4.2. Noncooperative Players' Optimization. Given the optimal CPs, which can be found in functional form, we can then state the following.

4.2.1. Players' Problem. Each VNF i wants to minimize, with respect to its workload vector $\mathbf{f}^i = \mathrm{col}[f_1^i, \ldots, f_N^i]$, a weighted (over its workload distribution) sum of the resources' costs, given the flow distribution of the others:

$$\begin{aligned} f^{i^*} &= \underset{f_1^i,\dots,f_N^i:\sum_{j=1}^N f_j^i=\lambda_i}{\operatorname{argmin}} J^i \\ &= \underset{f_1^i,\dots,f_N^i:\sum_{j=1}^N f_j^i=\lambda_i}{\operatorname{argmin}} \frac{1}{\lambda_i}\sum_{j=1}^{N} f_j^i c_j^*\left(f_j^i\right) \quad i = 1,\dots,S. \end{aligned} \tag{8}$$

In this formulation, the players' problem is a noncooperative game, of which we can seek a NE [31].

We examine the case of the application of CP1 first. Then, the cost of VNF i is given by

$$J^i = \frac{1}{\lambda_i}\sum_{j=1}^{N} f_j^i h_j \cdot \left(f_j\right)^2 = \frac{1}{\lambda_i}\sum_{j=1}^{N} f_j^i h_j \cdot \left(f_j^i + f_j^{-i}\right)^2, \tag{9}$$

where $f_j^{-i} = \sum_{k\neq i} f_j^k$ is the total flow from the other VNFs to node j.

The cost in (9) is convex in f_j^i and belongs to a category of cost functions previously investigated in the networking literature, without specific reference to energy efficiency [32, 33]. In particular, it is in the family of functions considered in [33], for which the NE Point (NEP) existence and uniqueness conditions of Rosen [34] have been shown to hold. Therefore, our players' problem admits a unique NEP. The latter can be found by considering the Karush-Kuhn-Tucker stationarity conditions with Lagrange multiplier ξ_i:

$$\begin{aligned} \xi_i &= \frac{\partial J^i}{\partial f_k^i}, \quad f_k^i > 0, \\ \xi_i &\le \frac{\partial J^i}{\partial f_k^i}, \quad f_k^i = 0, \\ & k = 1,\dots,N, \end{aligned} \tag{10}$$

from which, for $f_k^i > 0$,

$$\begin{aligned} \lambda_i\xi_i &= h_k\left(f_k^i + f_k^{-i}\right)^2 + 2h_k f_k^i\left(f_k^i + f_k^{-i}\right), \\ f_k^i &= -\frac{2}{3}f_k^{-i} \pm \frac{1}{3h_k}\sqrt{\left(h_k f_k^{-i}\right)^2 + 3h_k\lambda_i\xi_i}. \end{aligned} \tag{11}$$

Excluding the negative solution, the nonzero components are those with the smallest and equal partial derivatives that, in a subset $\mathcal{M} \subseteq \{1,\dots,N\}$, yield $\sum_{j\in\mathcal{M}} f_j^i = \lambda_i$; that is,

$$\sum_{j\in\mathcal{M}}\left[-\frac{2}{3}f_j^{-i} + \sqrt{4\left(h_j f_j^{-i}\right)^2 + 3h_j\lambda_i\xi_i}\right] = \lambda_i \tag{12}$$

from which ξ_i can be determined.

As regards form (7) of the optimal node cost, we can note that if we are restricted to $\underline{f}_j \le f_j \le \overline{f}_j$, $j = 1,\dots,N$ (a coupled convex constraint set), the conditions of Rosen still hold. If we allow the larger constraint set $0 < f_j < \mu_j^{\max}$, $j = 1,\dots,N$, the nodes' optimal cost functions are no longer of polynomial type. However, the composite function is still continuously differentiable (see the Appendix). Then, it would (i) satisfy the conditions for a convex game (the overall function composed of the three parts is convex), which guarantee the existence of a NEP [34, Th. 1], and (ii) possess equivalent properties to functions of Type A as defined in [32], which lead to uniqueness of the NEP.

5. Numerical Results

In order to evaluate our criterion, we present numerical results deriving from the application of the simplest Control Policy (CP1), which implies the solution of a completely quadratic problem (corresponding to costs as in (9) for the NEP). We compare the resulting allocations, power consumption, and average delays with those stemming from the application (on non-energy-aware nodes) of the algorithm for the minimization of the pure delay function that was developed in [35, Proposition 1]. This algorithm is considered a valid reference point in order to provide good validation of our criterion. The corresponding cost of VNF i is

$$J_T^i = \sum_{j=1}^{N} f_j^i T_j\left(f_j\right), \quad i = 1,\dots,S, \tag{13}$$

where

$$T_j\left(f_j\right) = \begin{cases} \dfrac{1}{\mu_j^{\max} - f_j}, & f_j < \mu_j^{\max}, \\ \infty, & f_j \ge \mu_j^{\max}. \end{cases} \tag{14}$$

The total demand is assumed to be less than the total operating capacity, as we have done with our CP2 in (7).

To make the comparison fair, we have to make sure that the maximum operating capacities $\mu_j^{\max}$, $j = 1,\dots,N$ (that are constant in (14)), are compatible with the quadratic behavior stemming from the application of CP1. To this aim, let ${}^{(\mathrm{LCP1})}\overline{f}_j$ denote the total input workload to node j corresponding to the NEP obtained by our problem; we then choose

$$\mu_j^{\max} = \vartheta_j\,{}^{(\mathrm{LCP1})}\overline{f}_j, \quad j = 1,\dots,N. \tag{15}$$

We indicate with ${}^{(T)}f_j$, ${}^{(T)}f_j^i$, $j = 1,\dots,N$, $i = 1,\dots,S$, the total node input workloads and those corresponding to VNF i, respectively, produced by the NEP deriving from the minimization of costs in (13). The average delays for the flow of player i under the two strategy profiles are obtained as

$$\begin{aligned}
D_T^i &= \frac{1}{\lambda_i}\sum_{j=1}^{N}\frac{{}^{(T)}f_j^i}{\mu_j^{\max} - {}^{(T)}f_j} \\
&= \frac{1}{\lambda_i}\sum_{j=1}^{N}\frac{{}^{(T)}f_j^i}{\vartheta_j {}^{(\mathrm{LCP1})}\overline{f}_j - {}^{(T)}f_j}, \\
D_{\mathrm{LCP1}}^i &= \frac{1}{\lambda_i}\sum_{j=1}^{N}\frac{{}^{(\mathrm{LCP1})}\overline{f}_j^i}{\vartheta_j {}^{(\mathrm{LCP1})}\overline{f} - {}^{(\mathrm{LCP1})}\overline{f}_j} \\
&= \frac{1}{\lambda_i}\sum_{j=1}^{N}\frac{{}^{(\mathrm{LCP1})}\overline{f}_j^i}{\left(\vartheta_j - 1\right){}^{(\mathrm{LCP1})}\overline{f}_j},
\end{aligned} \tag{16}$$

and the power consumption values are given by

$$\begin{aligned}
{}^{(T)}P_j &= K_j\left(\mu_j^{\max}\right)^3 = K_j\left(\vartheta_j {}^{(\mathrm{LCP1})}\overline{f}_j\right)^3, \\
{}^{(\mathrm{LCP1})}P_j &= K_j\left(\vartheta_j {}^{(\mathrm{LCP1})}\overline{f}_j\right)^3\left(\frac{\overline{f}_j}{\vartheta_j {}^{(\mathrm{LCP1})}\overline{f}_j}\right)^{1/\alpha_j} \\
&= K_j\left(\vartheta_j {}^{(\mathrm{LCP1})}\overline{f}_j\right)^3 \vartheta_j^{-1/\alpha_j}.
\end{aligned} \tag{17}$$

Our data for the comparison are organized as follows. We consider normalized input rates, so that $\lambda_i \le 1$, $i = 1, \dots, S$. We generate randomly S values corresponding to the VNFs' input rates from independent uniform distributions; then,

(1) we find the NEP of our quadratic game, by choosing the parameters K_j, α_j, $j = 1, \dots, N$, also from independent uniform distributions (with $1 \le K_j \le 10$, $2 \le \alpha_j \le 3$, resp.);

(2) by using the workload profiles obtained, we compute the values $\mu_j^{\max}$, $j = 1, \dots, N$, from (15) and find the NEP corresponding to costs in (13) and (14), by using the same input rates;

(3) we compute the corresponding average delays and power consumption values for the two cases, by using (16) and (17), respectively.

Steps (1)–(3) are repeated with a new configuration of random values of the input rates for a certain number of times; then, for both games, we average the values of the delays and power consumption per VNF, and of the total delay and power consumption averaged over the VNFs, over all trials. We compare the results obtained in this way for four different settings of VNFs and nodes; namely, $S = N = 3$; $S = 3$, $N = 5$; $S = 5$, $N = 3$; $S = 5$, $N = 5$. The rationale of repeating the experiments with random configurations is to explore a number of different possible settings and collect the results in a synthetic form.

For each VNFs-nodes setting, 30 random input configurations are generated to produce the final average values. In Figures 2–10, the labels EE and NEE refer to the energy-efficient case (quadratic cost) and to the non-energy-efficient one (see (13) and (14)), respectively. Instead, the label "Saving (%)" refers to the saving (in percentage) of the EE case compared to the NEE one. When it is negative, the EE case presents higher values than the NEE one. Finally, the label "User" refers to the VNF.

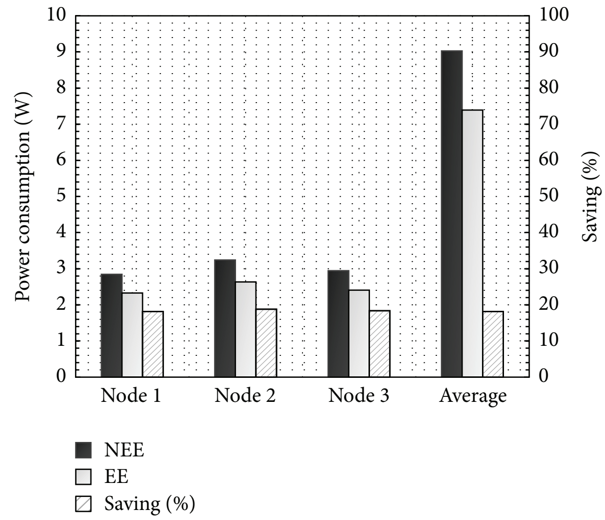

Figure 2: Power consumption with $S = N = 3$.

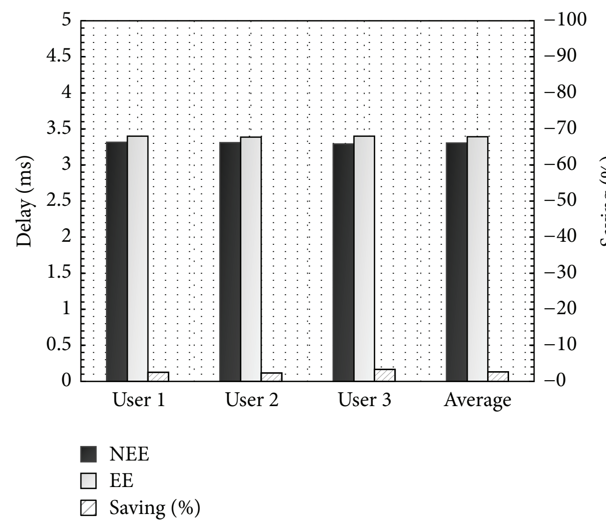

Figure 3: User (VNF) delays with $S = N = 3$.

The results are reported in Figures 2–10. The model implementation and the relative results have been developed with MATLAB® [36].

In general, the energy-efficient optimization tends to save energy at the expense of a relatively small increase in delay. Indeed, as shown in the figures, the saving is positive for the energy consumption and negative for delay. The energy saving is always higher than 18% in every case, while the delay increase is always lower than 4%.

However, it can be noted (by comparing, e.g., Figures 5 and 7) that configurations with lower load are penalized in the delay. This effect is caused by the behavior of the simple linear adaptation strategy, which maintains constant utilization, and

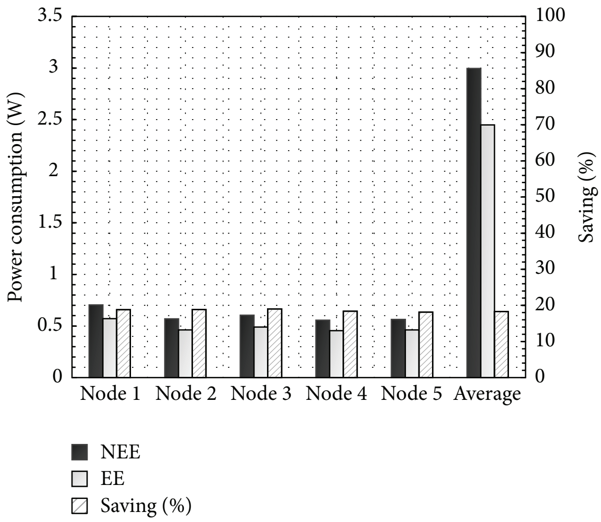

Figure 4: Power consumption with $S = 3$, $N = 5$.

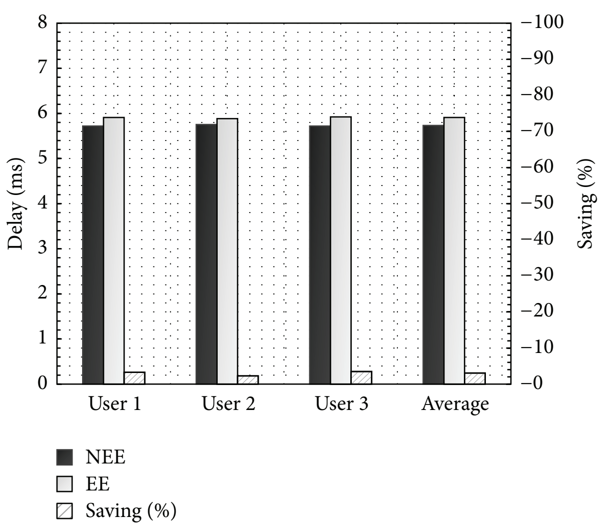

Figure 5: User (VNF) delays with $S = 3$, $N = 5$.

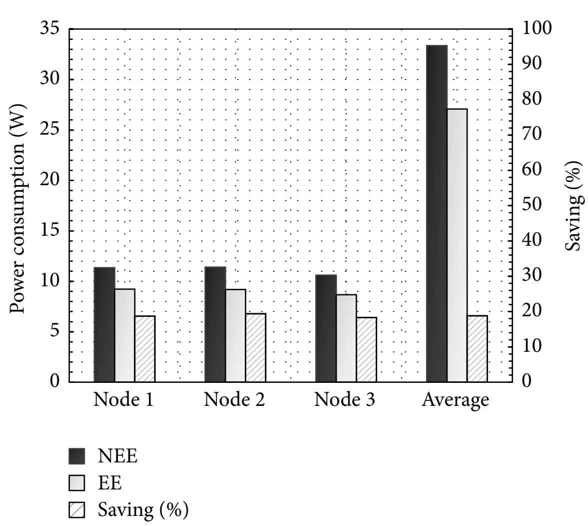

Figure 6: Power consumption with $S = 5$, $N = 3$.

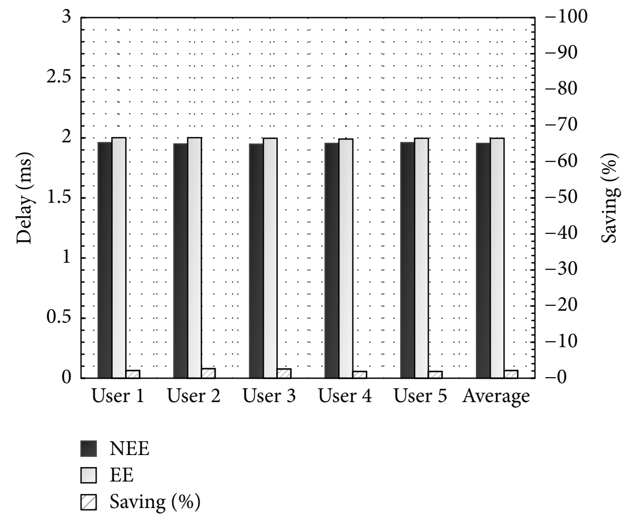

Figure 7: User (VNF) delays with $S = 5$, $N = 3$.

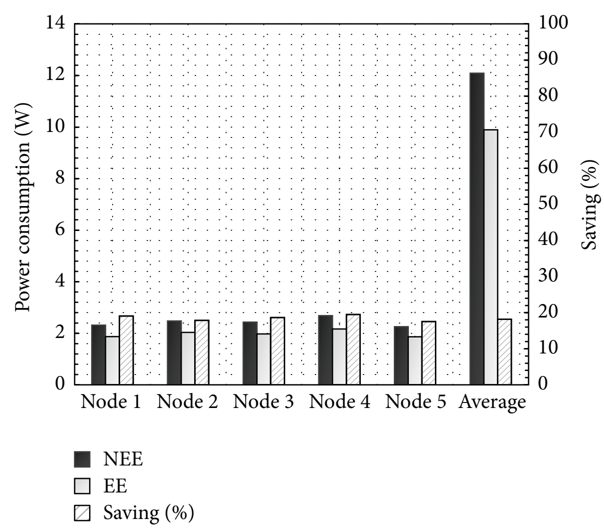

Figure 8: Power consumption with $S = 5$, $N = 5$.

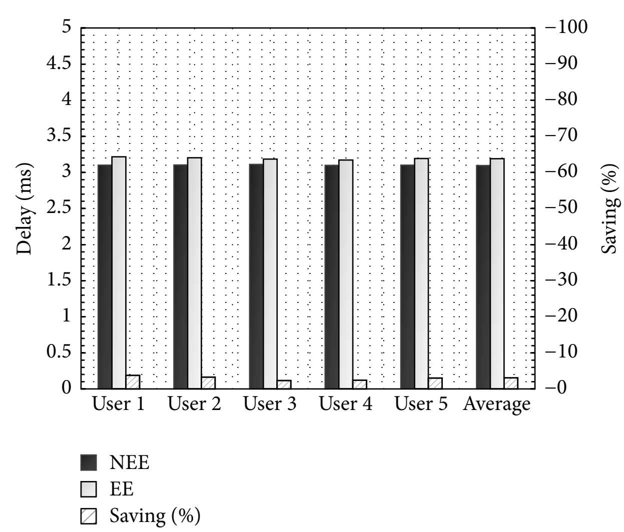

Figure 9: User (VNF) delays with $S = 5$, $N = 5$.

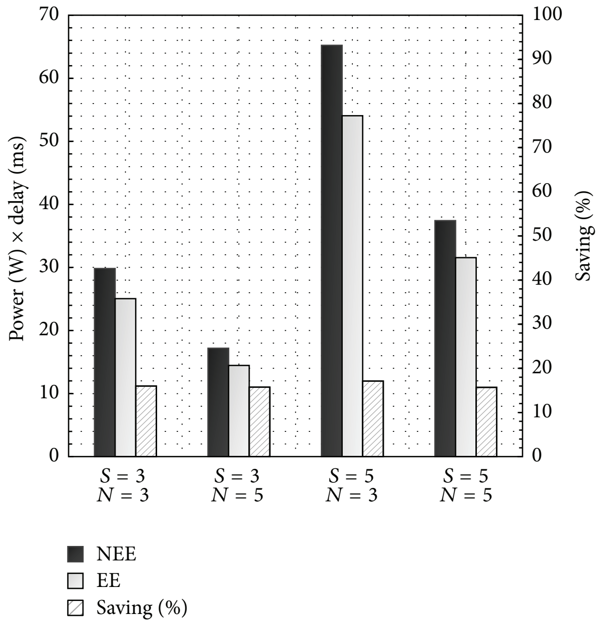

Figure 10: Power consumption and delay product of the various cases.

highlights the relevance of setting not only maximum, but also minimum values for the processing capacity (or for the node input workloads).

Another consideration arising from the results concerns the variation of the energy consumption, by changing the number of players and/or the nodes. For example, Figure 6 (case $S = 5$, $N = 3$) shows total power consumption significantly higher than the one shown in Figure 8 (case $S = 5$, $N = 5$). The reasons for this difference are due to the lower workload managed by the nodes of the case $S = 5$, $N = 5$ (greater number of nodes with an equal number of users) making it possible to exploit the AR mechanisms with significant energy saving.

Finally, Figure 10 shows the product of the power consumption and the delay for each studied case. As described in the previous sections, our criterion aims at minimizing this product. As can be easily seen, the saving resulting from the application of our policy is always above 10%.

6. Conclusions

We have examined the possible effect of introducing simple energy optimization strategies in physical network nodes performing services for a set of VNFs that request their computational power within an NFV environment. The VNFs' strategy profiles for resource sharing have been obtained from the solution of a noncooperative game, where the players are aware of the adaptive behavior of the resources, and aim at minimizing costs that reflect this behavior. We have compared the energy consumption and processing delay with those stemming from a game played with non-energy-aware nodes and VNFs. The results show that the effect on delay is almost negligible, whereas significant power saving can be obtained. In particular, the energy saving is always higher than 18% in every case, with a delay increase always lower than 4%.

From another point of view, it would also be of interest to investigate socially optimal solutions stemming from a Nash Bargaining Problem (NBP), along the lines of [37], by defining suitable utility functions for the players. Another interesting evolution of this work could be the application of additional constraints to the proposed solution such as, for example, the maximum delay that a flow can support.

Appendix

We check the maintenance of continuous differentiability at the point $\overline{f}_j = \mu_j^{\max}/\vartheta_j$. By noting that the derivative of the cost function of player i with respect to f_j^i has the form

$$\frac{\partial J^i}{\partial f_j^i} = c_j^*\left(f_j\right) + f_j^i c_j^{*\prime}\left(f_j\right), \tag{A.1}$$

it is immediate to note that we need to compare

$$\frac{\partial}{\partial f_j^i} K_j \frac{\left(f_j^i + f_j^{-i}\right)^{1/\alpha_j}\left(\mu_j^{\max}\right)^{3-1/\alpha_j}}{\mu_j^{\max} - f_j^i - f_j^{-i}}\Bigg|_{f_j^i=\mu_j^{\max}/\vartheta_j - f_j^{-i}}, \qquad \frac{\partial}{\partial f_j^i} h_j \cdot \left(f_j\right)^2\Bigg|_{f_j^i=\mu_j^{\max}/\vartheta_j - f_j^{-i}}. \tag{A.2}$$

We have

$$K_j \frac{\left(1/\alpha_j\right)\left(f_j\right)^{1/\alpha_j-1}\left(\mu_j^{\max}\right)^{3-1/\alpha_j}\left(\mu_j^{\max} - f_j\right) + \left(f_j\right)^{1/\alpha_j}\left(\mu_j^{\max}\right)^{3-1/\alpha_j}}{\left(\mu_j^{\max} - f_j\right)^2}\Bigg|_{f_j=\mu_j^{\max}/\vartheta_j}$$

$$= \frac{K_j\left(\mu_j^{\max}\right)^{3-1/\alpha_j}\left(\mu_j^{\max}/\vartheta_j\right)^{1/\alpha_j}\left[\left(1/\alpha_j\right)\left(\mu_j^{\max}/\vartheta_j\right)^{-1}\left(\mu_j^{\max} - \mu_j^{\max}/\vartheta_j\right) + 1\right]}{\left(\mu_j^{\max} - \mu_j^{\max}/\vartheta_j\right)^2}$$

$$= \frac{K_j\left(\mu_j^{\max}\right)^3\left(\vartheta_j\right)^{-1/\alpha_j}\left[\left(1/\alpha_j\right)\vartheta_j\left(1-1/\vartheta_j\right)+1\right]}{\left(\mu_j^{\max}\right)^2\left(1-1/\vartheta_j\right)^2} = \frac{K_j\mu_j^{\max}\left(\vartheta_j\right)^{-1/\alpha_j}\left(\vartheta_j/\alpha_j-1/\alpha_j+1\right)}{\left(\left(\vartheta_j-1\right)/\vartheta_j\right)^2},$$

$$2h_jf_j\big|_{f_j=\mu_j^{\max}/\vartheta_j} = 2K_j\frac{\left(\vartheta_j\right)^{3-1/\alpha_j}}{\vartheta_j-1}\cdot\frac{\mu_j^{\max}}{\vartheta_j}. \tag{A.3}$$

Then, let us check when

$$\begin{aligned}
&\frac{K_j\mu_j^{\max}\left(\vartheta_j\right)^{-1/\alpha_j}\left(\vartheta_j/\alpha_j-1/\alpha_j+1\right)}{\left(\left(\vartheta_j-1\right)/\vartheta_j\right)^2} \\
&\quad = 2K_j\frac{\left(\vartheta_j\right)^{3-1/\alpha_j}}{\vartheta_j-1}\cdot\frac{\mu_j^{\max}}{\vartheta_j}; \\
&\frac{\left(\vartheta_j/\alpha_j-1/\alpha_j+1\right)\cdot\left(\vartheta_j\right)^2}{\vartheta_j-1} = 2\left(\vartheta_j\right)^2; \\
&\frac{\vartheta_j}{\alpha_j}-\frac{1}{\alpha_j}+1 = 2\vartheta_j-2; \\
&\left(\frac{1}{\alpha_j}-2\right)\frac{3\alpha_j-1}{2\alpha_j-1}-\frac{1}{\alpha_j}+3 = 0; \\
&\left(\frac{1-2\alpha_j}{\alpha_j}\right)\frac{3\alpha_j-1}{2\alpha_j-1}-\frac{1}{\alpha_j}+3 = 0; \\
&\frac{1-3\alpha_j}{\alpha_j}-\frac{1}{\alpha_j}+3 = 0; \\
&1-3\alpha_j-1+3\alpha_j = 0.
\end{aligned} \tag{A.4}$$

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was partially supported by the H2020 European Projects ARCADIA (http://www.arcadia-framework.eu/) and INPUT (http://www.input-project.eu/), funded by the European Commission.

References

[1] C. Bianco, F. Cucchietti, and G. Griffa, "Energy consumption trends in the next generation access network—a telco perspective," in *Proceedings of the 29th International Telecommunication Energy Conference (INTELEC '07)*, pp. 737–742, Rome, Italy, September-October 2007.

[2] GeSI SMARTer 2020: The Role of ICT in Driving a Sustainable Future, December 2012, http://gesi.org/portfolio/report/72.

[3] A. Manzalini, "Future edge ICT networks," *IEEE COMSOC MMTC E-Letter*, vol. 7, no. 7, pp. 1–4, 2012.

[4] B. A. A. Nunes, M. Mendonca, X.-N. Nguyen, K. Obraczka, and T. Turletti, "A survey of software-defined networking: past, present, and future of programmable networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1617–1634, 2014.

[5] "Network functions virtualisation—introductory white paper," in *Proceedings of the SDN and OpenFlow World Congress*, Darmstadt, Germany, October 2012.

[6] R. Bolla, R. Bruschi, F. Davoli, and C. Lombardo, "Fine-grained energy-efficient consolidation in SDN networks and devices," *IEEE Transactions on Network and Service Management*, vol. 12, no. 2, pp. 132–145, 2015.

[7] R. Bolla, R. Bruschi, F. Davoli, and F. Cucchietti, "Energy efficiency in the future internet: a survey of existing approaches and trends in energy-aware fixed network infrastructures," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 2, pp. 223–244, 2011.

[8] I. Menache and A. Ozdaglar, *Network Games: Theory, Models, and Dynamics*, Morgan & Claypool Publishers, San Rafael, Calif, USA, 2011.

[9] R. Bolla, R. Bruschi, F. Davoli, and P. Lago, "Optimizing the power-delay product in energy-aware packet forwarding engines," in *Proceedings of the 24th Tyrrhenian International Workshop on Digital Communications—Green ICT (TIWDC '13)*, pp. 1–5, IEEE, Genoa, Italy, September 2013.

[10] A. Khan, A. Zugenmaier, D. Jurca, and W. Kellerer, "Network virtualization: a hypervisor for the internet?" *IEEE Communications Magazine*, vol. 50, no. 1, pp. 136–143, 2012.

[11] Q. Duan, Y. Yan, and A. V. Vasilakos, "A survey on service-oriented network virtualization toward convergence of networking and cloud computing," *IEEE Transactions on Network and Service Management*, vol. 9, no. 4, pp. 373–392, 2012.

[12] G. M. Roy, S. K. Saurabh, N. M. Upadhyay, and P. K. Gupta, "Creation of virtual node, virtual link and managing them in network virtualization," in *Proceedings of the World Congress on Information and Communication Technologies (WICT '11)*, pp. 738–742, IEEE, Mumbai, India, December 2011.

[13] A. Manzalini, R. Saracco, C. Buyukkoc et al., "Software-defined networks or future networks and services—main technical challenges and business implications," in *Proceedings of the IEEE Workshop SDN4FNS*, Trento, Italy, November 2013.

[14] M. Yu, Y. Yi, J. Rexford, and M. Chiang, "Rethinking virtual network embedding: substrate support for path splitting and migration," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 2, pp. 17–19, 2008.

[15] A. Jarray and A. Karmouch, "Decomposition approaches for virtual network embedding with one-shot node and link mapping," *IEEE/ACM Transactions on Networking*, vol. 23, no. 3, pp. 1012–1025, 2015.

[16] N. M. M. K. Chowdhury, M. R. Rahman, and R. Boutaba, "Virtual network embedding with coordinated node and link mapping," in *Proceedings of the 28th Conference on Computer Communications (IEEE INFOCOM '09)*, pp. 783–791, Rio de Janeiro, Brazil, April 2009.

[17] X. Cheng, S. Su, Z. Zhang et al., "Virtual network embedding through topology-aware node ranking," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 2, pp. 38–47, 2011.

[18] J. F. Botero, X. Hesselbach, M. Duelli, D. Schlosser, A. Fischer, and H. De Meer, "Energy efficient virtual network embedding," *IEEE Communications Letters*, vol. 16, no. 5, pp. 756–759, 2012.

[19] A. Leivadeas, C. Papagianni, and S. Papavassiliou, "Efficient resource mapping framework over networked clouds via iterated local search-based request partitioning," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 6, pp. 1077–1086, 2013.

[20] A. Fischer, J. F. Botero, M. T. Beck, H. de Meer, and X. Hesselbach, "Virtual network embedding: a survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 1888–1906, 2013.

[21] M. Schöller, M. Stiemerling, A. Ripke, and R. Bless, "Resilient deployment of virtual network functions," in *Proceedings of the 5th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT '13)*, pp. 208–214, Almaty, Kazakhstan, September 2013.

[22] A. Jarray and A. Karmouch, "VCG auction-based approach for efficient Virtual Network embedding," in *Proceedings of the IFIP/IEEE International Symposium on Integrated Network Management (IM '13)*, pp. 609–615, Ghent, Belbium, May 2013.

[23] A. Gember-Jacobson, R. Viswanathan, C. Prakash et al., "OpenNF: enabling innovation in network function control," in *Proceedings of the ACM Conference on Special Interest Group on Data Communication (SIGCOMM '14)*, pp. 163–174, ACM, Chicago, Ill, USA, August 2014.

[24] J. Antoniou and A. Pitsillides, *Game Theory in Communication Networks: Cooperative Resolution of Interactive Networking Scenarios*, CRC Press, Boca Raton, Fla, USA, 2013.

[25] M. S. Seddiki, M. Frikha, and Y.-Q. Song, "A non-cooperative game-theoretic framework for resource allocation in network virtualization," *Telecommunication Systems*, vol. 61, no. 2, pp. 209–219, 2016.

[26] L. Pavel, *Game Theory for Control of Optical Networks*, Springer, New York, NY, USA, 2012.

[27] E. Altman, T. Boulogne, R. El-Azouzi, T. Jiménez, and L. Wynter, "A survey on networking games in telecommunications," *Computers & Operations Research*, vol. 33, no. 2, pp. 286–311, 2006.

[28] S. Nedevschi, L. Popa, G. Iannaccone, S. Ratnasamy, and D. Wetherall, "Reducing network energy consumption via sleeping and rate-adaptation," in *Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation (NSDI '08)*, pp. 323–336, San Francisco, Calif, USA, April 2008.

[29] S. Henzler, *Power Management of Digital Circuits in Deep Sub-Micron CMOS Technologies*, vol. 25 of *Springer Series in Advanced Microelectronics*, Springer, Dordrecht, The Netherlands, 2007.

[30] K. Li, "Scheduling parallel tasks on multiprocessor computers with efficient power management," in *Proceedings of the IEEE International Symposium on Parallel & Distributed Processing, Workshops and PhD Forum (IPDPSW '10)*, pp. 1–8, IEEE, Atlanta, Ga, USA, April 2010.

[31] T. Başar, *Lecture Notes on Non-Cooperative Game Theory*, Hamilton Institute, Kildare, Ireland, 2010, http://www.hamilton.ie/ollie/Downloads/Game.pdf.

[32] A. Orda, R. Rom, and N. Shimkin, "Competitive routing in multiuser communication networks," *IEEE/ACM Transactions on Networking*, vol. 1, no. 5, pp. 510–521, 1993.

[33] E. Altman, T. Başar, T. Jiménez, and N. Shimkin, "Competitive routing in networks with polynomial costs," *IEEE Transactions on Automatic Control*, vol. 47, no. 1, pp. 92–96, 2002.

[34] J. B. Rosen, "Existence and uniqueness of equilibrium points for concave N-person games," *Econometrica*, vol. 33, no. 3, pp. 520–534, 1965.

[35] Y. A. Korilis, A. A. Lazar, and A. Orda, "Architecting noncooperative networks," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, pp. 1241–1251, 1995.

[36] MATLAB®, The Language of Technical Computing, http://www.mathworks.com/products/matlab.

[37] H. Yaïche, R. R. Mazumdar, and C. Rosenberg, "A game theoretic framework for bandwidth allocation and pricing in broadband networks," *IEEE/ACM Transactions on Networking*, vol. 8, no. 5, pp. 667–678, 2000.