Towards Efficient Offloading in Fog/Edge Computing by Approximating Effect of Externalities

Vladimir Marbukh
National Institute of Standards & Technology
100 Bureau Drive, Stop 8910
Gaithersburg, MD 20899-8910
marbukh@nist.gov

Abstract—**Fog/Edge computing is an emerging architecture, which extends the Cloud computing paradigm to the edge of the network, enabling new applications and services, including Internet of Things (IoT). End devices have certain computation tasks, which can be completed either locally on the end device or remotely on the cloud via computation offloading. Due to mobility and highly dynamic nature of end users, the access is wireless with severe physical limitations on the access network capacity to sustain stringent latency requirements for streaming and real-time applications for large number of end users over wide-spread geographical area. Assuming that access network has cellular architecture and is interference limited, we quantify system performance by the aggregate utility. Optimization of the system computing and communication resources requires accounting for externalities due to interference created by the individual offloading transmissions. While each base station can account for the intracell externalities, exact accounting for the intercell externalities in a large-scale network is not feasible due to prohibitive exchange of the "microscopic" information on the intercell interference by each end user. We propose a "macroscopic" approximation for the intercell externalities, which significantly reduces information exchange, and thus allows for decentralized near-maximization of the aggregate system utility.**

Keywords-Fog/Edge computing; offloading; utility; resource management; externalities; approximation.

I. INTRODUCTION

Fog/Edge computing is an emerging architecture that moves computation, communication, and storage closer to the end users [1]-[2]. The emergence of Fog/Edge computing is driven by advent of the Internet of Things (IoT) and enabled by a variety of powerful end-user, network edge, and access devices with embedded artificial intelligence and 5G communication capabilities. These devices include smartphones, tablets, smart home appliances, small cellular base stations, edge routers, traffic control devices, connected vehicles, smart meters, and energy controllers in a smart power grid, smart building controllers, industrial control systems, drones, and industrial and consumer robots.

While "Cloud paradigm" assumes moving computing, control, and data storage into the centralized cloud, "Fog/Edge paradigm" relies on balancing centralized and local computing, storage, and network management. However, finding the "right balance" is a challenging problem due to "very large scale" of IoT and strong externalities, i.e., side effects of local resource management decisions. Recently developed Network Utility Maximization (NUM) based cross-layer network optimization [3] allows for distributed "social optimization" by pricing externalities in some classes of networks. However, large scale and the highly dynamic nature and heterogeneity of resources in the emerging Fog/Edge computing paradigm make real-time exchange of implied costs of offloading decisions by individual end users "too pricy" with respect to communication resources. This paper proposes extension of NUM [3] to make it applicable to Fog/Edge computing. This extension allows for approximate pricing of long-range externalities at the cost of some loss in aggregate system utility.

Figure 1 shows a simplified view of a Fog/edge computing architecture [1]-[2].

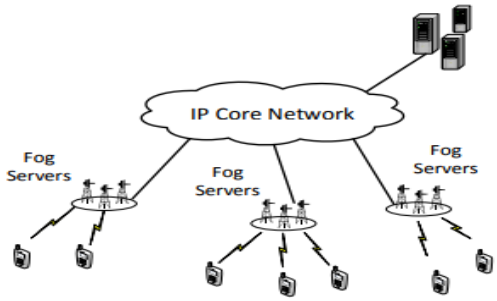

Figure 1. Fog/edge computing architecture.

Each mobile device has certain computation tasks, which can be completed either locally on the mobile device or remotely on the cloud via computation offloading. Due to mobility and highly dynamic nature of end users, the access is wireless with severe physical limitations on the access network capacity to sustain stringent latency requirements for streaming and real-time applications for large number of mobile nodes over wide-spread geographical area.

Assuming the wireless access network to be interference limited, we quantify system performance by the aggregate user utility. Optimization of the system computing and communication resources requires accounting for externalities due to interference created by the individual offloading transmissions. While each base station can account for the intracell externalities, exact accounting for the intercell

externalities in a large-scale network is not feasible due to prohibitive exchange of the "microscopic" information on the intercell interference by each mobile. We propose a "macroscopic" approximation for the intercell externalities, which significantly reduces information exchange, and thus allows for decentralized near-maximization of the aggregate system utility.

The paper is organized as follows. Section II describes communication model of interference limited cellular wireless access in Fog/edge network. Extending results [4], we show that the feasible region for end user transmission rates can be *exactly* quantified in much lower dimension macro-parameters, which are interference at the base stations. Section III quantifies user and system performance by user and aggregate utility respectively. Selfish user performance optimization, while being distributed, results in inefficiency with respect to system performance operating point. We propose approximate system performance optimization, which allows for a distributed implementation at the cost of *approximate* pricing for inter-cell externalities. Finally, the conclusion briefly summarizes and outlines directions for future research.

II. Communication Model

Subsection A describes interference limited cellular network. Given user offloading rates, mobile transmission powers satisfy a non-negative system of linear algebraic equations of dimension equal to the number of end users. The solution to this "microscopic" system can be exactly recovered from a "macroscopic" linear system for interferences at the base stations. Dimension of this system, equal to the number of base stations, is typically much lower than the number of end users. Subsection B gives concise characterization of achievable transmission rates in terms of Perron-Frobenius eigenvalue to the non-negative matrix of the macroscopic system.

A. Transmission Powers

Consider a cellular network comprised of N cells. Cell $n=1,..,N$ covers K_n users. We identify each user by a pair (n,k), where n identifies the cell and $k=1,..,K_n$ identifies the number of end users in the cell. User (n,k) Signal to Interference plus noise Ratio is

$$SINR_{nk} = p_{nk}\xi_{nk}^n / (I_{nk}+\sigma_n^2), \tag{1}$$

where p_{nk} is user (n,k) transmission power, ξ_{sm}^n is propagation gain from user (s,m) to base station n, σ_n^2 is exogenous interference at base station n, and endogenous interference experienced by user (n,k) at base station n from other end users is

$$I_{nk} = \sum\nolimits_{(m,s)\neq(n,k)} p_{ms}\xi_{ms}^n . \tag{2}$$

We assume that user (n,k) transmission rate is an increasing function of the Signal to SINR (1):

$$r_{nk} = \varphi_{nk}\left(\frac{p_{nk}\xi_{nk}^n}{\sum\nolimits_{(m,s)\neq(n,k)} p_{ms}\xi_{ms}^n + \sigma_n^2}\right), \tag{3}$$

where $\varphi_{nk}(0)=0$

Two examples of rate function (3) are Shannon capacity:

$$r_{nk} = W\log_2(1+SINR_{nk}), \tag{4}$$

where W is the wireless bandwidth, and threshold capacity:

$$r_{ns} = \begin{cases} r_{ns}^* & if \quad SINR_{ns} > SINR_{ns}^* \\ 0 & otherwise \end{cases}. \tag{5}$$

Threshold-based capacity (5) is an extreme case of a realistic sigmoid rate function of a wireless channel.

Introducing total interference at the base station n

$$I_n = \sum_{(m,s)} p_{ms}\xi_{ms}^n \tag{6}$$

and inverting (3), we obtain the following expression for the transmission powers:

$$p_{ms} = [\varphi_{ms}^{-1}(r_{ms})/\xi_{ms}^m](I_m+\sigma_m^2 - p_{ms}\xi_{ms}^m), \tag{7}$$

where $\varphi_{nk}^{-1}(.)$ is inverse of increasing function $\varphi_{nk}(.)$. Relations (6)-(7) form a closed system of linear equations of transmission powers by users. Dimension of this system is equal to the number of end users $\sum_{n=1}^N K_n$, which may be quite high. Following [4], in the rest of this subsection we show that transmission powers p_{nk} can be recovered from solving a linear system of much lower dimension N.

Equation (7) yields

$$p_{ms} = \frac{\varphi_{ms}^{-1}(r_{ms})}{[1+\varphi_{ms}^{-1}(r_{ms})]\xi_{ms}^m}(I_m+\sigma_m^2). \tag{8}$$

Multiplying both sides of (8) by ξ_{ms}^n and then summing over (m,s), we obtain the following equations for interference at base stations I_n, $n=1,..,N$:

$$I_n = \sum_m (I_m+\sigma_m^2)\sum_s\left(\frac{\varphi_{ms}^{-1}(r_{ms})}{1+\varphi_{ms}^{-1}(r_{ms})}\frac{\xi_{ms}^n}{\xi_{ms}^m}\right). \tag{9}$$

Moving term containing I_n from the right-hand to the left-hand side, we obtain the following closed system of N linear equations for interferences at base stations I_n, $n=1,..,N$:

$$I_n = \frac{1}{1-\gamma_n}\left[\left(\sum_{m\neq n} I_m \sum_s \gamma_{ms}\frac{\xi_{ms}^n}{\xi_{ms}^m}\right)+\left(\sum_m \sigma_m^2 \sum_s \gamma_{ms}\frac{\xi_{ms}^n}{\xi_{ms}^m}\right)\right] \tag{10}$$

where

$$\gamma_{ns}(r_{ns}) = \frac{\varphi_{ns}^{-1}(r_{ns})}{1+\varphi_{ns}^{-1}(r_{ns})}, \tag{11}$$

$$\gamma_n(r) = \sum\nolimits_s \gamma_{ns}(r_{ns}), \tag{12}$$

and $r=(r_{ns})$ is vector of transmission rates. After solving linear system (10)-(12) for interference I_n, user transmission powers p_{ns} can be recovered with explicit expressions (8).

In the case of a single cell, equation (10) yields explicit expression for interference at the base station

$$I = \frac{\gamma(r)}{1-\gamma(r)}\sigma^2, \tag{13}$$

which yields end user transmission powers

$$p_s = \frac{\sigma^2}{1-\gamma(r)} \frac{\varphi_s^{-1}(r_s)}{[1+\varphi_s^{-1}(r_s)]\xi_s}, \tag{14}$$

where

$$\gamma(r) = S - \sum_s \frac{1}{1+\varphi_s^{-1}(r_s)}. \tag{15}$$

B. Transmission Rates

According to (13)-(15), the capacity region of a single-cell system is given by

$$\sum_s \frac{1}{1+\varphi_s^{-1}(r_s)} > S-1. \tag{16}$$

Given Shannon capacity (4),

$$\varphi_{ns}^{-1}(r) = 2^{r_{ns}/W} - 1, \tag{17}$$

and thus sustainability condition (16) takes the following form:

$$S < 1 + \sum_s 2^{-r_s/W}. \tag{18}$$

Given threshold-based capacity (5),

$$\varphi_{ns}^{-1}(r) = \begin{cases} SINR_{ns}^* & if \quad r = r_{ns}^* \\ 0 & if \quad r = 0 \end{cases}, \tag{19}$$

and thus, sustainability condition (16) takes the following form:

$$S < 1 + \sum_s \frac{1}{1+SINR_s^*}. \tag{20}$$

For a multi-cell system,

$$\gamma_n(r) < 1, \; n = 1,..,N, \tag{21}$$

with $\gamma_n(r)$ given by (12), is generally a necessary but not sufficient condition for sustainability of rates $r = (r_{nk})$. Assuming (21) are satisfied, in the rest of this subsection we concisely define the capacity region of a multi-cell system

$$R = \left\{ r : \middle| \exists p \ge 0 : r_{nk} = \varphi_{nk}\left(\frac{p_{nk}\xi_{nk}^n}{\sum_{(m,s)\neq(n,k)} p_{ms}\xi_{ms}^n + \sigma_n^2} \right) \right\} \tag{22}$$

Our analysis is based on observation that equations (10)-(12) form a linear system:

$$I = A(r)I + b(r), \tag{23}$$

where matrix $A(r) = [A_{nm}(r)]_{n,m=1}^N$ has non-negative components

$$A_{nm}(r) = \frac{1}{1-\gamma_n(r)} \sum_s \gamma_{ms}(r_{ms}) \frac{\xi_{ms}^n}{\xi_{ms}^m}, \tag{24}$$

if $n \neq m$, and $A_{nm} = 0$, and column vector $b = (b_n)_{n=1}^N$ has positive components

$$b_n(r) = \frac{1}{1-\gamma_n(r)} \sum_m \sigma_m^2 \sum_s \gamma_{ms}(r_{ms}) \frac{\xi_{ms}^n}{\xi_{ms}^m}. \tag{25}$$

According to (8), transmission rates $r = (r_{nk})$ can be realized with finite transmission powers if and only if system (23) has non-negative solution $I_n \ge 0$, $n = 1,..,N$, i.e.,

$$R = \{ r : \middle| \exists I \ge 0 : \; I = A(r)I + b(r) \}. \tag{26}$$

It is known [5] that capacity region (26) can be characterized in terms of Perron-Frobenius eigenvalue of matrix $A(r)$ with components (24), $\Gamma = \Gamma(r)$:

$$\Gamma(r) < 1 \tag{27}$$

complemented with conditions (21). Conditions (21), (27) provide a concise Perron-Frobenius characterization of system capacity region (22), which is open and generally non-convex.

We conclude this subsection with the following brief notes. First, convexification of the capacity region (21), (27) can be achieved with time sharing between different transmission power vectors [5]. Second, approximations and bounds on the Perron-Frobenius eigenvalue $\Gamma(r)$ immediately lead to corresponding approximations and bounds on the system capacity region (27)-(28). For example, it is known [6] that

$$\Gamma(r) \le \max_n \sum\nolimits_{m\neq n} A_{nm}(r), \tag{28}$$

and thus, condition

$$\max_n \sum_{m\neq n} \frac{1}{1-\sum_s \varphi_{ns}^{-1}(r_{ns})} \sum_s \varphi_{ms}^{-1}(r_{ms})(\xi_{ms}^n / \xi_{ms}^m) \le 1 \tag{29}$$

guarantees (27). Third, in a case of a single cell, $N = 1$, conditions (21), (29) are necessary and sufficient for sustainability of rates $r = (r_{1k})$.

III. Performance Optimization

System operation inside a capacity region but close to its boundary results in a high level of interference and thus requires a high level of transmission power. For mobile end users, transmission power is inversely related to battery life, so mobile users should balance their preferences for transmission rate against prolonging battery life. Subsection A introduces user utility functions that quantify this tradeoff. Subsection B formulates a game-theoretic framework for selfish user utility maximization and a social-optimization framework for aggregate utility maximization. Subsection C proposes and discusses distributed implementation of social near optimization, which is based on approximate pricing of inter-cell externalities.

A. User Utility

The tradeoff between user (n,k) competing preferences for transmission rate r and power p can be quantified by utility function $u_{nk}(r,p)$, which is increasing in $r \ge 0$ and decreasing in $p \ge 0$. We assume utilities to be non-negative: $u_{nk}(r,p) \ge 0$, and moreover: $u_{nk}(r,p) = 0$, $u_{nk}(r,p) \downarrow 0$ as $p \uparrow \infty$.

For file transfer and delay-sensitive applications, Figures 2a and 2b sketch utility $u_{nk}(r,p)$ as a function of transmission rate r for different transmission powers $p = p_1, p_2, p_3$, where $p_1 < p_2 < p_3$. Given p, utility $u(r,p)$ is increasing in r since if not for communication constraints, due to superior computational capability, centralized computation is preferable to local processing. A specific form of end-user utility as a function of transmission rate in a case of computation offloading has been discussed in literature,

e.g., see [7]. Here we note only that in the case of file transfer (Figure 2a) this function is concave, and in case of delay-sensitive applications (Figure 2b) this function has sigmoid shape.

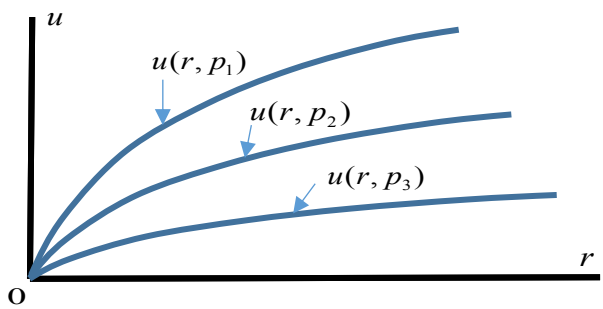

Figure 2a. File transfer: $p_1 < p_2 < p_3$.

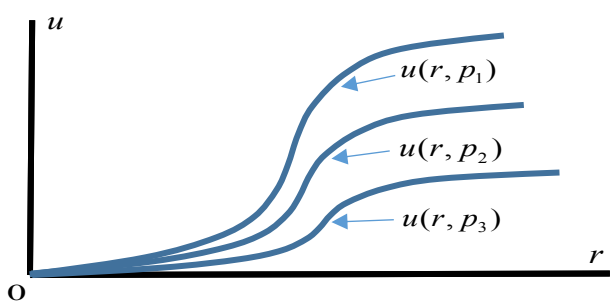

Figure 2b. Latency-sensitive applications: $p_1 < p_2 < p_3$.

Figure 3 sketches user utility $u(r,p)$ as a decreasing function of transmission power p (due to user preference for conservation of battery energy) for different transmission rates $r = r_1, r_2, r_3$, where $r_1 < r_2 < r_3$.

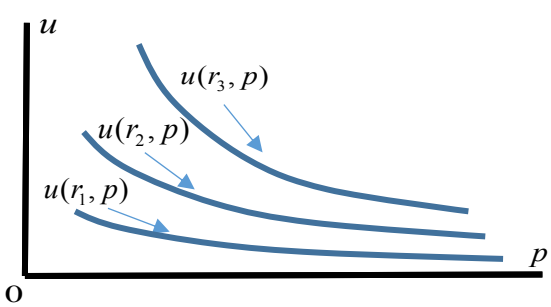

Figure 3. Utility vs. transmission power p for $r_1 < r_2 < r_3$.

Substituting expression (3) for the transmission rate r_{nk} in user (n,k) utility $u_{nk}(r_{nk}, p_{nk})$, we obtain user utility as a function of the transmission power:

$$U_{nk}(p_{nk}|q_{nk}) = u_{nk}[\varphi_{nk}(p_{nk}/q_{nk}), p_{nk}], \tag{30}$$

where the "effective interference" is

$$q_{nk} = (I_{nk} + \sigma_n^2)/\xi_{nk}^n. \tag{31}$$

Figure 4, which sketches utility (30) in a practically typical case of a sigmoid rate function (i.e., delay-sensitive application), demonstrates presence on user utility from two opposing effects of increasing transmission power p.

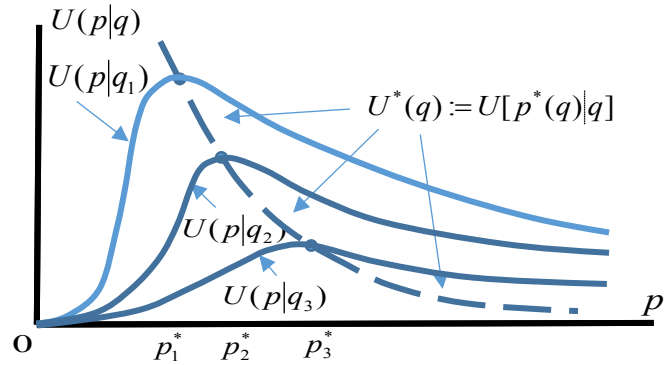

Figure 4. Utility $U(p|q)$ vs. p for $q = q_1 < q_2 < q_3$.

The beneficial effect due to increasing transmission rate dominates for sufficiently small transmission power, and the detrimental effect due to increasing energy consumption dominates for sufficiently large transmission power.

B. Selfish vs. Social Optimization

Given interference I_{nk} experienced by user (n,k) at the receiver, it is rational for user (n,k) to choose transmission power p_{nk}^*, which maximizes utility (30):

$$p_{nk}^* = \psi_{nk}(q_{nk}) := \arg\max_{p \ge 0} U_{nk}(p|q). \tag{32}$$

Figure 4 shows this internal user optimization and the corresponding optimal utility $U^*(q) := U[\psi(q)|q]$ in a typical case when transmission power p_{nk}^* is unique solution to

$$\partial U_{nk}(p_{nk}|I_{nk})/\partial p_{nk} = 0. \tag{33}$$

Since user (n,k) utility depends not only on this user transmission power but also on the transmission powers of other users through interference experienced at the base station n, given by (2), selfish optimization (32)-(33) can be naturally formalized as a non-cooperative game among users, and system equilibrium can be naturally associated with a pure Nash equilibrium in such a game. In the rest of this subsection we derive fixed-point equations for this equilibrium.

Combining (2), (6), and (30)-(32) we can express the result of selfish optimization (32) as a function of total interference at base station n, I_n, given by (6) as follows:

$$p_{nk} = \psi_{nk}\left(\frac{I_n + \sigma_n^2}{\xi_{nk}^n} - p_{nk}\right). \tag{34}$$

Since $\psi_{nk}(x)$ is a decreasing function (see Figure 4), equation (34) has a unique solution with respect to p_{nk}:

$$p_{nk} = \pi_{nk}(I_{nk}). \tag{35}$$

Substituting (35) into (6) we obtain a closed system of fixed-point equations for interference at the base stations:

$$I_n = \sum_{(m,s)} \xi_{ms}^n \pi_{ms}(I_m). \tag{36}$$

Asymptotically stable equilibria of system (36) represent Nash equilibria of the corresponding selfish optimization game. Neither existence nor uniqueness of finite equilibria are

guaranteed. A case of increasing to infinity fixed-points (36) describes a situation of "power warfare" when users keep raising their transmission powers without limit [ ].

As an example, consider a step-wise rate function (5), when $\psi_{nk}(q) = SINR^*_{nk} q$, and thus equation (34) takes the following form:

$$p_{nk} = [(I_n + \sigma_n^2)/\xi^n_{nk} - p_{nk}]SINR^*_{nk}. \quad (37)$$

Equation (37) yields

$$p_{nk} = [(I_n + \sigma_n^2)/\xi^n_{nk}]\omega_{nk} \quad (38)$$

and thus system (36) takes the following form:

$$I_n = \frac{\sum_{m\neq n} I_m \sum_s \frac{\xi^n_{ms}}{\xi^m_{ms}}\omega_{ms} + \sum_{(m,s)} \sigma_m^2 \frac{\xi^n_{ms}}{\xi^m_{ms}}\omega_{ms}}{1 - \sum_s \omega_{ns}}, \quad (39)$$

where $\omega_{nk} = SINR^*_{nk}/(1 + SINR^*_{nk})$. The necessary and sufficient conditions for convergence of the linear fixed-point system (39) are a combination of inequalities $\sum_s \omega_{ns} < 1$, $n = 1,..,N$ and the condition that the Perron-Frobenius eigenvalue of a non-negative matrix of linear system (39) is upper bounded by unity. If at least one of these conditions is not satisfied, transmission powers of all users grow to infinity, which results in zero utility for all users. This demonstrates a possibility for high inefficiency of selfish user optimization (32) with respect to aggregate system utility

$$U_\Sigma(p|I) = \sum_{(n,k)} U_{nk}(p_{nk}|I_{nk}). \quad (40)$$

Social optimization framework attempts to maximize the aggregate system utility:

$$U_\Sigma(p^{opt}|I^{opt}) := \max_{p_{nk}\geq 0} U_\Sigma(p|I) \quad (41)$$

subject to constraints (2). Generally, this optimization problem is not convex, and thus solution based on Lagrange multipliers produces only a local optimum. The corresponding Lagrangian is

$$L = \sum_{(n,k)} U_{nk}(p_{nk}|I_{nk}) - \sum_{(n,k)} \lambda_{nk}\left(\sum_{(m,s)\neq(n,k)} p_{ms}\xi^n_{ms} - I_{nk}\right). \quad (42)$$

The first order optimality conditions with respect to p_{nk} and I_{nk} are as follows:

$$\partial U_{nk}(p_{nk}|I_{nk})/\partial p_{nk} = \sum_{(m,s)\neq(n,k)} \lambda_{ms}\xi^m_{nk}, \quad (43)$$

$$\partial U_{nk}(p_{nk}|I_{nk})/\partial I_{nk} = -\lambda_{nk}, \quad (44)$$

Conditions (43)-(44) should be supplemented with (2).

Inefficiency of selfish user optimization (32) can be quantified as Price of Anarchy [8]:

$$PoA = U_\Sigma(p^{opt}|I^{opt})/U_\Sigma(p^*|I^*) \geq 1. \quad (45)$$

Extreme inefficiency $PoA = \infty$ occurs in a case of "power warfare," when selfish users raise transmission power to infinity, and thus $U_\Sigma(p^*|I^*) = 0$. Comparing (33) with (43) reveals that the source inefficiency in selfish user optimization (32) is externalities, since the right-hand side in (43) quantifies the effect of externalities. Next subsection proposes an approximate evaluation of this effect, which allows for a distributed implementation, and may result in near optimization of system performance.

C. Distributed System Performance Optimization

Selfish user (n,k) optimization, assuming marginal cost π_{nk} on transmission power p, is as follows

$$\max_{p\geq 0}[U_{nk}(p|I_{nk}) - \pi_{nk}p]. \quad (46)$$

First order optimality condition for problem (46) coincides with first order social optimality condition (43) if

$$\pi_{nk} = \sum_s \lambda_{ns}\xi^n_{nk} + \sum_{m\neq n} \Lambda_m \xi^m_{nk}, \quad (47)$$

where according to (44), Lagrange multiplier λ_{ns} quantifies user (n,s) willingness to pay for marginal reduction in experienced interference, and thus $\Lambda_m = \sum_s \lambda_{ms}$ is the aggregate willingness to pay by all users in cell n.

Assuming for now that user (n,k) is aware of current implied cost (47), consider the following adaptation:

$$\dot{p}_{nk} = \partial U_{nk}(p_{nk}|I_{nk})/\partial p_{nk} - \pi_{nk}, \quad (48)$$

which results in the following evolution of aggregate utility (41)

$$\frac{dU_\Sigma(p|I)}{dt} = \sum_{(n,k)} \left(\frac{\partial U_{nk}(p_{nk}|I_{nk})}{\partial p_{nk}} + \sum_{(m,s)\neq(n,k)} \frac{\partial U_{ms}(p_{ms}|I_{ms})}{\partial I_{ms}} \frac{\partial I_{ms}}{\partial p_{nk}} \right) \dot{p}_{nk}. \quad (49)$$

According to (2), $\partial I_{ms}/\partial p_{nk} = \xi^m_{nk}$, and thus

$$\frac{dU_\Sigma(p|I)}{dt} = \sum_{(n,k)} \left(\frac{\partial U_{nk}(p_{nk}|I_{nk})}{\partial p_{nk}} - \sum_{(m,s)\neq(n,k)} \lambda_{ms}\xi^m_{nk} \right) \dot{p}_{nk} = \sum_{(n,k)} \left(\frac{\partial U_{nk}(p_{nk}|I_{nk})}{\partial p_{nk}} - \pi_{nk} \right)^2 \geq 0. \quad (50)$$

According to (50), $-U_\Sigma(p|I)$ is a Lyapunov function for adaptation (48), and thus converges to "closest" local maximum of the aggregate system utility.

Practical implementation of adaptation (48) in large-scale systems with highly dynamic users is not feasible due to unacceptably high overhead required for near real-time updates of the implied costs (47). Component $\sum_s \lambda_{ns}\xi^n_{nk}$ of cost (47), which quantifies the cost of intra-cell externalities, can be effectively evaluated since each base station n can collect willingness to pay λ_{nk} and estimate propagation gain ξ^n_{nk} from all users (n,k) in a cell. However, component $\sum_{m\neq n} \Lambda_m \xi^m_{nk}$ of cost (47), which quantifies implied cost of inter-cell externalities, cannot be effectively evaluated due to uncertain propagation gains ξ^m_{nk} from each user (n,k) to each base station $m \neq n$. A natural approach to overcome this obstacle is to approximate or bound implied cost of inter-cell externalities. In the rest of this subsection we make some specific suggestions to this effect, assuming that propagation

gains between all base stations n and m, ξ_n^m are available since base stations are more stationary than end users.

Probably, the simplest approximation assumes that path gain from user (n,k) to a "non-native" base station $m \neq n$ can be approximated by path gain from the local base station n to base station m: $\xi_{nk}^m \approx \xi_n^m$, $n \neq m$, and thus,

$$\pi_{nk} \approx \tilde{\pi}_{nk} := \sum_s \lambda_{ns}\xi_{nk}^n + \sum_{m\neq n} \Lambda_m \xi_n^m . \quad (51)$$

Adaptation (49), where π_{nk} is replaced with $\tilde{\pi}_{nk}$,

$$\dot{\tilde{p}}_{nk} = \partial U_{nk}(\tilde{p}_{nk}|I_{nk}) / \partial \tilde{p}_{nk} - \tilde{\pi}_{nk} , \quad (52)$$

produces an approximation for user transmission powers: $p_{nk} \approx \tilde{p}_{nk}$. This approximation is accurate in extreme cases of very high and very low wireless signal attenuation. In the former case, the effect of inter-cell interference is negligible as compared to intra-cell interference. In the latter case, interference at a base station is a sum of interferences from users belonging to a large number of cells and is not very sensitive to user locations within each cell. This suggests applicability of approximation (51)-(52) in intermediate cases of wireless signal propagation.

We demonstrate a possibility of bounding implied cost of inter-cell externalities:

$$\breve{\pi}_{nk} \leq \pi_{nk} \leq \hat{\pi}_{nk} \quad (53)$$

on an example free propagation model. However, note that the bounds proposed below are valid for other practically important propagation models. The low bound in (53)

$$\breve{\pi}_{nk} = \sum_s \lambda_{ns}\xi_{nk}^n + \xi_{nk}^n \sum_{m\neq n} \Lambda_m \xi_n^m \quad (54)$$

follows from the following bound on the propagation gain $\xi_{nk}^m \geq \xi_{nk}^n \xi_n^m$. The upper bound in (53) is as follows:

$$\hat{\pi}_{nk} = \sum_s \lambda_{ns}\xi_{nk}^n + \sum_{m\neq n} \Lambda_m \min_{l\in J_n} \xi_l^m , \quad (55)$$

where J_n is the set of cells neighboring cell n. Assuming that each base station n is aware of set J_n is natural since a system should be able to deal with handoffs. Combining bounds (53)-(55) with adaptation (48) we obtain the following differential inequality:

$$\frac{\partial U_{nk}(p_{nk}|I_{nk})}{\partial p_{nk}} - \hat{\pi}_{nk} \leq \dot{p}_{nk} \leq \frac{\partial U_{nk}(p_{nk}|I_{nk})}{\partial p_{nk}} - \breve{\pi}_{nk} , \quad (56)$$

which bounds user transmission powers: $\breve{p}_{nk} \leq p_{nk} \leq \hat{p}_{nk}$.

IV. Conclusion and Future Research

Efficient balancing between centralized and local computing, storage, and network management in emerging Fog/Edge computing infrastructure requires accounting for externalities created by offloading decisions made by individual end users. However, overhead associated with pricing "long-range" externalities may be unacceptably high. To mitigate this difficulty, we have proposed an Extended Network Utility Maximization (ENUM) framework, which reduces this overhead by allowing for approximate pricing of "long-range" externalities at the cost of some loss in system performance.

Numerous issues deserve future investigation, including accuracy, performance, practical applicability of proposed bounds and approximations, as well as developing new boundaries and approximations. The most fundamental issue yet to be addressed is optimization of system performance with respect to accuracy of the "near real-time" implied cost of the offloading decisions by individual users. On the one hand, making implied costs more accurate by increasing frequency of updates, allows users to better align their individual offloading decisions with overall system performance. On the other hand, frequent updates of implied costs reduce physically limited wireless bandwidth available for user data. One may envision a scheme that adapts update frequency of implied costs to their rate of change.

References

[1] F. Bonomi, R. Milito, P. Natarajan, J. Zhu, "Fog computing: A platform for Internet of Things and analytics" in Big Data and Internet of Things: A Roadmap for Smart Environments, Cham, Springer, pp. 169-186, 2014.

[2] M. Chiang and T. Zhang, "Fog and IoT: an overview of research opportunities," IEEE Internet of Things Journal, Vol. 3, No. 6, 2016.

[3] M. Chiang, S. H. Low, A. R. Calderbank, J. C. Doyle, "Layering as optimization decomposition: A mathematical theory of network architectures", Proc. IEEE, vol. 95, no. 1, pp. 255-312, Jan. 2007.

[4] S.V. Hanly, "Congestion measures in DS-CDMA networks," IEEE Trans. on Communications, Vol.: 47, Issue: 3, Mar 1999.

[5] W. Yu and J. Yuan, "Joint source coding, routing and resource allocation for wireless sensor networks," IEEE ICC 2005.

[6] Meyer, Matrix Analysis and Applied Linear Algebra, SIAM, 2000.

[7] X. Chen, "Decentralized computation offloading game for mobile cloud computing," IEEE Transactions on Parallel and Distributed Systems, 2014.

[8] M. Liu and Y. Wu, "Spectrum sharing as congestion games", 46th Annual Communication Control and Computing Allerton Conference, pp. 1146-1153, 2008.