Towards Fog Network Utility Maximization (FoNUM) for Managing Fog Computing Resources

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

Abstract—**Fog 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 users have certain computation tasks, which can be completed either locally on the end device or remotely on an accessible Fog node 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. Also, computationally and storage intensive tasks can be offloaded to centralized cloud over IP core network at the cost of higher communication delay. Efficiency of offloading critically depends not only on the wireless access to Fog nodes, but also on availability of the computing, storage, control, and networking resources at the Fog nodes. This position paper proposes Fog Network Utility Maximization (FoNUM) for balancing end user preferences for various Fog services with mobile end user preferences for conserving battery energy to prolong battery lives. We suggest that approximate, pricing-based, distributed solution to FoNUM can be obtained by employing soft handoff, which allows peripheral devices to connect to several "close" Fog nodes and then customize the offloading depending on the specific task resource requirements and resource availability at the Fog nodes.**

Keywords-Fog network; utility maximization; resource management; pricing; approximation, soft handoff.

I. Introduction

Fog computing is an emerging architecture that moves computation, communication, and storage closer to the end users [1]-[2]. The emergence of Fog 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, industrial and consumer robots, etc.

While "Cloud paradigm" assumes moving computing, control, and data storage into the centralized cloud, "Fog 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," highly dynamic nature of IoT, and strong externalities, i.e., side effects of local resource management decisions. Recently developed Network Utility Maximization (NUM) [3] demonstrated feasibility of distributed cross-layer network optimization by pricing externalities for balancing user demands for communication bandwidth.

This position paper proposes Fog Network Utility Maximization (FoNUM) as an extension of NUM [3]. FoNUM balances end user preferences for various services with mobile user preferences for conserving battery energy to prolong battery lives. We argue that approximate, pricing-based, distributed solution to FoNUM can be obtained by employing soft handoff, which allows peripheral devices to connect to several "close" Fog nodes and then customize the job offloading depending on the job resource requirements and resource availability at the Fog nodes.

Figure 1 shows a "bird view" architecture of Fog computing architecture [1]-[2].

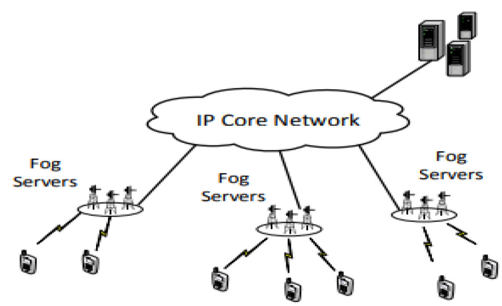

Figure 1. Fog computing architecture.

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 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 Fog node 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 end users. We suggest that approximate, pricing-based, distributed solution to FoNUM can be obtained by employing soft handoff, which allows peripheral devices to connect to several "close" Fog nodes and then customize the offloading depending on the specific task resource requirements and resource availability at the Fog nodes.

The paper is organized as follows. Section II describes interference limited wireless access network to the Fog network and service model at the Fog nodes, which extends model [4]. Given upper limits on the interference at the Fog nodes, the Fog capacity region is the intersection of the communication capacity region of the access network to the Fog nodes and capacity regions for computing services at the Fog nodes. Extending results [5], Section III shows that the communication capacity region of the access network to the Fog nodes can be effectively evaluated using Perron-Frobenius theory. Section IV describes Fog Network Utility Maximization (FoNUM) framework as argues that efficient resource allocation in Fog network can be achieved with distributed, pricing-based solution to FoNUM. Finally, conclusion briefly summarizes and outlines directions of future research.

II. Communication and Computing Models

Subsection A describes interference limited wireless access to Fog nodes, used by end users for offloading. Subsection B describes computing service model at Fog nodes. Given upper limits on the interference at the Fog nodes, these models yield the feasible region for the offloading rates.

A. Access Communication Network to Fog Nodes

Consider Fog network with S peripheral users and N Fog nodes. User $i = 1,..,M$ Signal to Interference plus noise Ratio at Fog node $n = 1,..,N$ is

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

where p_{in} is user i transmission power to node n, ξ_{in} is propagation gain from user i to Fog node n, σ_n^2 is exogenous interference at node n, and interference experienced by user i at node n from other users $j \neq i$ is

$$I_{in} = \sum\nolimits_{j \neq i} \sum\nolimits_{k=1}^{N} p_{jk}\xi_{jn}\ . \tag{2}$$

Expression (2) assumes that each user coordinates her transmissions to different nodes, thus these transmissions do not interfere with each other. However, different nodes do not coordinate transmissions with each other.

We assume that user i transmission rate is an increasing function of the Signal to Interference plus noise Ratio (1):

$$r_{in} = \varphi_{in}\left(\frac{p_{in}\xi_{in}}{I_{in} + \sigma_n^2}\right), \tag{3}$$

where $\varphi_{in}(0) = 0$. Two examples of rate function (3) are Shannon capacity:

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

where W is the wireless bandwidth, and threshold capacity:

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

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

Given the interference level at the Fog nodes

$$I_n = \sum\nolimits_{i=1}^{M} \sum\nolimits_{k=1}^{N} p_{ik}\xi_{in}\ , \tag{6}$$

equations (2)-(3) yield end user i transmission power needed for offloading rate to Fog node n:

$$p_{in} = \frac{\varphi_{in}^{-1}(r_{in})}{[1 + \varphi_{in}^{-1}(r_{in})]\xi_{in}}(I_n + \sigma_n^2)\,, \tag{7}$$

where $\varphi_{in}^{-1}(.)$ is inverse of increasing function $\varphi_{in}(.)$. Thus, given interference levels at the Fog nodes (6), the feasible region for the offloading rates is as follows:

$$\sum_{i=1}^{I} \sum_{k=1}^{N} \frac{\varphi_{ik}^{-1}(r_{ik})}{[1 + \varphi_{ik}^{-1}(r_{ik})]\xi_{in}}(I_k + \sigma_k^2) \le I_n\ . \tag{8}$$

Note that despite region (8) is not necessarily convex, it can be shown [6] that convexification can be achieved with time sharing between different transmission power vectors.

B. Computing Services at Fog Nodes

We assume that each Fog node has up to J types of resources in addition to the communication resources, which can be requested by an offloaded task. These additional resources may include computing resources, memory, network management resources, etc. Let B_{nj} be amount of resource $j = 1,..,J$ at Fog node $n = 1,..,N$. We assume that each end user $i = 1,..,M$ can offload up to S classes of tasks. Task offloading to a Fog node consumes task class specific mixture of resources at this node. We model the amount of consumed resources by assuming that offloading of a type $s = 1,..,S$ task at rate $r > 0$ consumes amount rc_j^s of type j resource at a Fog node.

Assuming that system can direct the offloading to different fog nodes, let r_{in}^s be user i transmission rate to node n required to sustain class s jobs generated by this peripheral. The total offloading rate by peripheral i to node n is

$$r_{in} = \sum\nolimits_{s} r_{in}^{s}\ . \tag{9}$$

Assuming that resource j capacity at node n is C_{jn}, node n can sustain demand for this resource if and only if

$$\sum\nolimits_{s} c_j^s \sum\nolimits_{i} r_{in}^s \le B_{jn}\ . \tag{10}$$

In Figure 2, region $0, R_{in}^s, R_{jn}^s, 0$ represents convexified feasible region due to communication constrains (8), region $0, A, AB, B, 0$ represents capacity region due to limited resources at the Fog nodes (10). Intersection of these two regions $F = 0, B, b', R_{jn}^s, 0$ represents the feasible region for the Fog network, given interference levels at the Fog nodes.

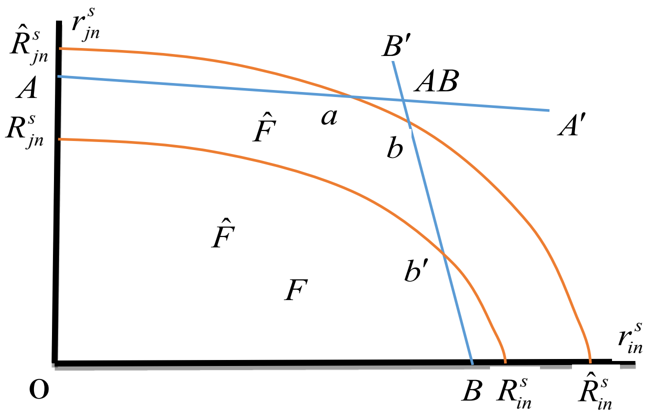

Figure 2. Feasible and capacity regions for offloading rates

III. Capacity Region

While inequalities (8) determine feasible user offloading rates for given interference levels at the Fog nodes, this section derives communication capacity of a Fog network, i.e., region of sustainable offloading rates. Subsection A demonstrates that solution to "microscopic" system (6)-(7) can be exactly recovered from "macroscopic" linear system for interferences at the Fog nodes. Dimension of this system, equal to the number of Fog nodes N, is typically much lower than number of the end users M: $N << M$. 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. Dimension Reduction

Multiplying both sides of (7) by ξ_{ik}, then summing over (i,n), and taking into account (6), we obtain the following equations for interference at base stations I_k, $k = 1,..,N$:

$$I_k = \sum_n (I_n + \sigma_n^2) \sum_i \left(\frac{\phi_{in}^{-1}(r_{in})}{1+\phi_{in}^{-1}(r_{in})} \frac{\xi_{ik}}{\xi_{in}} \right). \tag{11}$$

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

$$I_n = \frac{1}{1-\gamma_n} \left[\left(\sum_{k \neq n} I_k \sum_i \gamma_{ik} \frac{\xi_{in}}{\xi_{ik}} \right) + \left(\sum_k \sigma_k^2 \sum_i \gamma_{ik} \frac{\xi_{in}}{\xi_{ik}} \right) \right], \tag{12}$$

where

$$\gamma_{ik}(r_{ik}) = \frac{\varphi_{ik}^{-1}(r_{ik})}{1+\varphi_{ik}^{-1}(r_{ik})}, \tag{13}$$

$$\gamma_k(r) = \sum_i \gamma_{ik}(r_{ik}), \tag{14}$$

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

In a case of a single cell, (12)-(13) yield explicit expression for interference at the base station

$$I = \frac{\gamma(r)}{1-\gamma(r)} \sigma^2, \tag{15}$$

and end user transmission powers

$$p_i = \frac{\sigma^2}{1-\gamma(r)} \frac{\varphi_i^{-1}(r_s)}{[1+\varphi_i^{-1}(r_i)]\xi_i}, \tag{16}$$

where

$$\gamma(r) = I - \sum_i \frac{1}{1+\varphi_i^{-1}(r_i)}. \tag{17}$$

B. Perron-Frobenius Characterization

According to (16)-(17), capacity region of a single-cell system is given by

$$\sum_i \frac{1}{1+\varphi_i^{-1}(r_i)} > M-1. \tag{18}$$

In a case of Shannon capacity (4),

$$\phi_{ni}^{-1}(r) = 2^{r_{ni}/W} - 1, \tag{19}$$

and thus sustainability condition (18) takes the following form:

$$M < 1 + \sum_i 2^{-r_s/W}. \tag{20}$$

In a case of threshold-based capacity (5),

$$\varphi_{in}^{-1}(r) = \begin{cases} SINR_{in}^* & if \quad r = r_{in}^* \\ 0 & if \quad r = 0 \end{cases}, \tag{21}$$

and thus, sustainability condition (18) takes the following form:

$$M < 1 + \sum_i \frac{1}{1+SINR_i^*}. \tag{22}$$

For a multi-cell system,

$$\gamma_n(r) < 1, \; n = 1,..,N, \tag{23}$$

where $\gamma_n(r)$ are given by (13)-(14), is generally a necessary but not sufficient condition for sustainability of rates $r = (r_{in})$. Assuming conditions (23) are satisfied, in the rest of this subsection we concisely define the Fog communication capacity region in a general case of multiple Fog nodes:

$$\left\{ r : \left| \exists p \geq 0 : r_{nk} = \phi_{nk} \left(\frac{p_{nk}\xi_{nk}^n}{\sum_{(m,s)\neq(n,k)} p_{ms}\xi_{ms}^n + \sigma_n^2} \right) \right. \right\}. \tag{24}$$

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

$$I = A(r)I + b(r), \tag{25}$$

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

$$A_{nk}(r) = \frac{1}{1-\gamma_n(r)} \sum_i \gamma_{ik}(r_{ik}) \frac{\xi_{in}}{\xi_{ik}}, \tag{26}$$

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

$$b_n(r) = \frac{1}{1-\gamma_n(r)} \sum_k \sigma_k^2 \sum_i \gamma_{ik}(r_{ik}) \frac{\xi_{in}}{\xi_{ik}}. \tag{27}$$

According to (7), transmission rates $r = (r_{in})$ can be realized with finite transmission powers if and only if system (25)-(27) has non-negative solution $I_n \geq 0$, $n = 1,..,N$, i.e.,

$$\hat{R} = \{r : \left|\exists I \geq 0 : \ I = A(r)I + b(r)\right\}. \tag{28}$$

It is known [6] that capacity region (28) can be characterized in terms of Perron-Frobenius eigenvalue of matrix $A(r)$ with components (26), $\Gamma = \Gamma(r)$:

$$\Gamma\left(\sum\nolimits_s r_{in}^s\right) < 1 \tag{29}$$

complemented with conditions (23). Conditions (23), (29) provide a concise Perron-Frobenius characterization of system communication capacity region (24), which is open and generally non-convex. Convexification of the capacity region (23), (29) can be achieved with time sharing between different transmission power vectors [6]. In Figure 2 region $0, R_{in}^s, a, b, R_{jn}^s, 0$ represents this convexified communication capacity region. Intersection of the convexified communication capacity region and resource capacity region at the Fog nodes $0, A, AB, B, 0$ represents the overall Fog capacity region $\hat{F}$.

We conclude this subsection by noting that approximations and bounds on the Perron-Frobenius eigenvalue $\Gamma(r)$ immediately lead to the corresponding approximations and bounds on the Fog communication capacity region (24). For example, it is known [6] that

$$\Gamma(r) \leq \max_n \sum\nolimits_{k \neq n} A_{nk}(r), \tag{30}$$

and thus, condition

$$\max_n \frac{1}{1-\gamma_n(r)} \sum_{k \neq n} \sum_i \gamma_{ik}(r_{ik}) \frac{\xi_{in}}{\xi_{ik}} < 1 \tag{31}$$

supplemented with (23) guarantee (29), and thus determine low bound on the Fog communication capacity region (24). Also note that in a case of a single cell $N = 1$, conditions (23) and (31) the Fog communication capacity region (24).

IV. Fog Network Utility Maximization (FONUM)

System operation inside capacity region close to its boundary results in high level of interference and thus requires high level of transmission powers. Since for mobile end users, transmission power is inversely related to battery life expectancy, mobile users should balance their preferences for the offloading rate on the one hand and prolonging the battery life on the other hand. Subsection A introduces user utility functions which quantify this tradeoff. Assuming that overall Fog performance is characterized by the aggregate utility, subsection B demonstrates inefficiency of selfish user utility maximization. This inefficiency is due to strong negative externalities of the offloading decisions by individual end users. Subsection C discusses possible distributed implementation of Fog Network Utility Maximization (FoNUM), which is based on approximate pricing of these externalities.

A. User and System Performance Criteria

We assume that end user i preference for class s service can be quantified by increasing utility function $v_{is}(r_i^s)$, where $v_{is}(0) = 0$. Specific form of end user utility as a function of transmission rate in a case of computation offloading have been discussed in literature, e.g., see [8]. Here we only note that in a case of services which do not require minimum bandwidth, e.g., file transfer, this function is concave: $v = v_1(r)$, as shown in Figure 3. In a case of real-time, e.g., streaming, services, which require minimum bandwidth, this function has a sigmoid shape: $v = v_2(r)$, as shown in Figure 3.

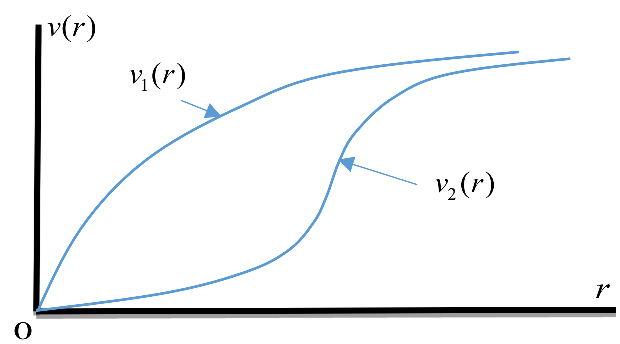

Figure 3. User utility of offloading

We assume that the aggregate utility of user i offloading services $s = 1,..,S$ at rates $r_i^s := (r_{is}, s = 1,..,S)$ is additive:

$$v_i^{\Sigma}(r_i^s) = \sum\nolimits_{s=1}^{S} {}_i^s(r_i^s). \tag{32}$$

Mobile end user i balances preference for high offloading rate with incentive for prolonging the battery life which translates to lower transmission power p_i. We model this balance by including penalty $f_i(p_i)$ in user i utility as follows:

$$u_i(r_i^s, p_i) = \sum\nolimits_s v_{is}(r_i^s) - f_i(p_i), \tag{33}$$

where $f_i(p_i)$ is an increasing and convex function, $f_i(0) = 0$ and $f_i(p_i) \uparrow \infty$ as $p_i \uparrow \infty$. Penalty function $f_i(p_i)$ is sketched in Figure 4.

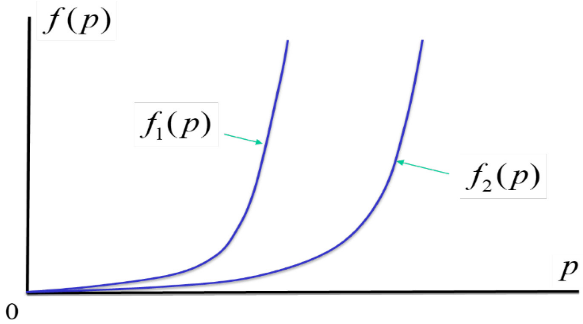

Figure 4. Penalty due to battery depleting

Function $f_1(p)$ describes mobile user with less remaining battery energy than function $f_2(p)$ describes.

B. *Inefficiency of Selfish Optimization*

In selfish optimization, each peripheral user $i \in I$ selfishly makes offloading decisions in an attempt to maximize hers individual utility (33), given interference levels at the Fog nodes I_n. Substituting expression (3) for the offloading rate r_{in} in user i utility (33), we obtain the following expression for the user utility i as a function of the transmission power, given interference levels at the Fog nodes I_n:

$$U_i(p_{in}^s|I_n) = \sum_s v_{is}\left[\sum_n \varphi_{in}\left(\frac{p_{in}^s \xi_{in}}{I_n - p_{in}^s \xi_{in} + \sigma_n^2}\right)\right] - \\ -f_i\left(\sum_n \sum_s p_{in}^s\right) \tag{34}$$

Thus, user i selfish optimization takes form of utility (34) maximization over transmission powers $p_{in}^s, s = 1,..,S$, $n = 1,..,N$:

$$\max_{(p_{in}^s)} U_i(p_{in}^s|I_n). \tag{35}$$

subject to resource capacity constraints (10).

Optimization (35) in a case of inactive capacity constraints (10) is shown in Figure 5 in a typical case of sigmoid rate function (3).

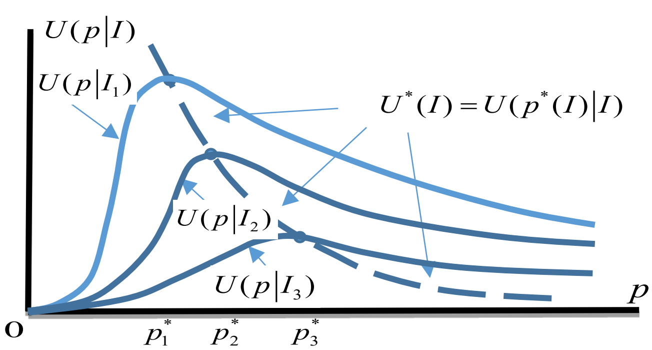

Figure 5. Utility: offloading vs. battery depleting, $I_1 < I_2 < I_3$

Increase in the user transmission power benefits user utility for sufficiently small transmission power due to increase in the offloading rate. However, for large transmission power, detrimental effect due to high battery energy expenditure dominates. Given interference level I, the optimal transmission power $p^*(I)$ is a decreasing function of I.

Since interference levels at the Fog nodes I_n depend on transmission powers by all interfering end users:

$$I_n(p) = \sum_{i=1}^M \sum_{k=1}^N \xi_{in} \sum_{s=1}^S p_{in}^s, \tag{36}$$

selfish user optimization can be naturally interpreted within game-theoretic framework. If fixed-point equations (6), (35) converge, the corresponding selfish equilibrium is a pure equilibrium of the corresponding game. Since $U(p|I) \downarrow 0$ as $p \uparrow 0$, system avoids "power warfare," i.e., unlimited increase in the transmission powers by end users competing for wireless bandwidth. Still, selfish equilibrium may be highly inefficient since "more transmission power challenged" end users with penalty function $f_1(p)$ in Figure 4 may be completely starved from offloading by "less transmission power challenged" end users with penalty function $f_2(p)$ in Figure 4.

Assuming that overall performance of Fog network is characterized by the aggregate utility

$$U_\Sigma[p|I(p)] = \sum_i U_i[p_{in}^s|I_n(p)], \tag{37}$$

a natural goal for Fog resource allocation is maximization of this aggregate utility:

$$U_\Sigma^{\max} = \max_{p \geq 0} U_\Sigma[p|I(p)] \tag{38}$$

subject to resource capacity constraints (10).

We call optimization problem (38)-(10) a Fog Network Utility Maximization (FoNUM) and quantify inefficiency of selfish optimization by the corresponding Price of Anarchy (PoA):

$$PoA^* = U_\Sigma^{\max} / U_\Sigma^* \geq 1, \tag{39}$$

where U_Σ^* is the aggregate utility (37) at the selfish equilibrium. The inefficiency of selfish optimization is due to not accounting for externalities due to interference at the Fog nodes, which may eliminate some users from offloading. Since exact accounting for the externalities in a large-scale Fog network is not feasible, in the next subsection we propose some approximations.

C. *FoNUM: Towards Distributed Solution through Pricing*

Following NUM framework, FoNUM accounts for externalities by imposing social costs on the users for the created externalities. The corresponding user i net utility is:

$$w_i(r_{in}^s, p_{in}^s) = u_i(r_{in}^s, p_i) - \sum_{n,s} r_{in}^s \sum_j g_{nj} c_j^s - q_n \sum_{n,s} p_{in}^s, \tag{40}$$

where g_{nj} is the marginal social cost of using of an unit of resource j capacity at Fog node n, and q_n is the marginal social cost of a unit of interference at Fog node n.

User $i \in I$ maximization of its net utility (40):

$$\max_{(p_{in}^s \ge 0)} w_i \left[\varphi_{in} \left(\frac{p_{in}^s \xi_{in}}{I_n - p_{in}^s \xi_{in} + \sigma_n^2} \right), p_{in}^s \right] \quad (41)$$

yield solution, which depends on the costs g_{nj} and q_i. The problem is finding the socially optimal costs, such that individual user net utility maximization (41) yields solution to FoNUM (37)-(38). In the rest of this subsection, we suggest that these socially optimal costs can be determined adaptively as an implementation of the demand/supply principle: cost increases (decreases) as resource supply decreases (increases).

Following [10], we consider algorithm which proceeds in discrete time $t = 0,1,2,..$. We assume that each Fog node $n = 1,..,N$ measures the aggregate demand on the type $j = 1,..,J$ resource at this node at time t

$$R_{jn}(t) = \sum_s c_j^s \sum_i r_{in}^s(t) \quad (42)$$

and evaluates the excess over available capacity B_{jn}: $R_{jn}(t) - B_{jn}$. Following demand/supply law for resource j at node n, the marginal social cost of using of an unit of resource j capacity at Fog node n evolves as follows:

$$g_{nj}(t+1) = \left[g_{nj}(t) + h_t (R_{jn}(t) - B_{jn}) \right]^+, \quad (43)$$

where $[x]^+ = \max(x, 0)$ and some sufficiently small $h_t > 0$. We also assume that each Fog node $n = 1,..,N$ estimates interference at this node at time t, $I_n(t)$, and evaluates the excessive interference over predetermined ermined target level $\widehat{I}_n$, $I_n(t) - \widehat{I}_n$. Following demand/supply law for the wireless bandwidth at node n, the marginal cost of a unit of interference at Fog node n evolves as follows:

$$q_n(t+1) = \left[q_n(t) + h_t (I_n(t) - \widehat{I}_n) \right]^+. \quad (44)$$

After costs (43)-(44) are communicated to users, these users determine Fog nodes to connect to and the transmission powers by solving their individual optimization problems (40)-(41).

The first issue with this algorithm is convergence for $h_t \downarrow 0$ as $t \to \infty$. The convergence can be expected based on convergence of similar algorithms in simpler situations [10]. However, for a Fog network, relevance of convergence is questionable due to arrivals/departures and mobility of end users. More relevant is an adequate ability to combine near optimal performance under steady scenario with adaptability under reasonable dynamic scenarios for sufficiently small but fixed $h_t = h > 0$. The second issue is selection of upper limits of interference levels at the Fog nodes $\widehat{I}_n$, $n = 1,..,N$ in (44). In accordance with FoNUM framework, this selection should be able to approximate solution to optimization problem (37)-(38). The third major issue is practically implementable strategy of updating end users on the costs (43)-(44) since only users expected to initiate soft handoff should be updated on the real-time costs (43)-(44) of the resources and interference at the "neighboring" Fog nodes. We outline possible ways to address these issues in the next section.

V. Conclusion and Future Research

Efficient balancing of the centralized and local computing, storage, and network management in emerging Fog computing infrastructure requires accounting for the externalities created by offloading decisions made by individual end users. This position paper suggests Fog Network Utility Maximization (FoNUM) for balancing various tradeoffs in Fog networks for each end user as well as across different users. FoNUM extends conventional Network Utility Maximization (NUM), which solves this problem in a case of communication resources.

We plan give recommendations on parameter h_t in (43)-(44) under realistic simulation scenarios. Selecting "near-optimal" target interference levels at the Fog nodes $\widehat{I}_n$, $n = 1,..,N$ as well as updating end users on the costs (43)-(44) requires evaluation of the long-range intracell externalities in near real time. Physical limitations on the wireless bandwidth create inherent tradeoffs between accuracy of the Fog management information and Fog ability to handle payload. One may envision a scheme which adapts frequency of updates of the implied costs according to the importance and rate of change. We plan investigation of such scheme by employing methodology of learning algorithms for soft handoff with recently emerged concept of the Age of Information (AoI) [11].

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] V. Marbukh, "Towards Efficient Offloading in Fog/Edge Computing by Approximating Effect of Externalities," 2nd Workshop on Integrating Edge Computing, Caching, and Offloading in Next Generation Network, IEEE Infocom, San Francisco, 2018.

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

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

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

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

[9] X. Lin, N.B. Shroff, and R. Srikant, A Tutorial on Cross-Layer Optimization in Wireless Networks, IEEE Journal on Selected Areas in Communications, Vol. 24, No. 8, August 2006.

[10] M. Costa, M. Codreanu, and A. Ephremides, "On the age of information in status update systems with packet management," IEEE Trans. Inf. Theory, vol. 62, no. 4, April 2016, 1897–1910.