Noname manuscript No.
(will be inserted by the editor)

Allais' Paradox and Resource Allocation in Telecommunication Networks

Philippe Ezran · Yoram Haddad · Mérouane Debbah

Received: date / Accepted: date

Abstract The question of resource allocation arises whenever demand exceeds supply. The common approach is to optimize the network efficiency while maintaining some fairness among the users. While resource allocation policies use various definitions for network efficiency and fairness, most of them are based on maximization of a utility function. The mathematical formalism underlying these approaches is the same as the mathematical formalism used in the Bernoulli model in finance, where a player is supposed to maximize his expected utility function. This model is disproved by Allais' paradox, which provides examples of rational behaviors which cannot be described by the maximization of any utility function. By transposing this paradox to telecommunication networks for the purpose of resource allocation, we build examples of rational operators whose optimal choice cannot be described by the maximization of any utility function. By optimizing a trade-off between network efficiency and fairness, we propose a model similar to the risk-return trade-off optimization in finance.

Keywords Resource allocation · Allais' paradox · Fairness · Network utility

Philippe Ezran
Jerusalem College of Technology, Israel
CentraleSupélec, France
Tel.: +972-52-3047817
E-mail: philippe.ezran@polytechnique.org

Yoram Haddad
Jerusalem College of Technology, Israel

Mérouane Debbah
CentraleSupélec, France

1 Introduction

The mobile telecommunication sector is presently enjoying a tremendous growth, and this trend is expected to continue in the foreseen future. According to an IDATE report [24], global mobile subscription should grow from 5,328 million in 2010 to 9,684 million in 2020 (+81.8%), while total worldwide mobile traffic should grow from 3.8 Ebyte in 2010 to 127 Ebyte in 2020 (x33).

This tremendous growth will inevitably cause saturation and congestion problems. In spite of the numerous solutions which are considered in order to meet the growing demand, there is no doubt that the question of resource allocation will become a key issue in the next years, especially in the scope of the introduction of 5G networks.

The question of resource allocation arises whenever demand exceeds supply. This issue goes far beyond the field of telecommunications and is extensively studied in economics and sociology. In telecommunication network management, the question arises especially in throughput allocation and base-station allocation in multi-tier networks [7], [17].

Some approaches favor the search of balance between energy consumption and throughput revenue [18], [5]. In addition, many methods have been proposed in order to improve energy efficiency [22], [16].

Other approaches introduce a pricing in order to optimize both energy consumption and throughput revenue [10], [15].

Resource allocation is also mentioned as a technique for maximization of network lifetime, where network lifetime is defined as the total amount of time during which the network is capable of maintaining its full functionality and/or of achieving particular objectives during its operation [30].

The common feature of all these approaches is that the question of resource allocation results from the fact that supply is limited. The objective is to optimize in some way the network efficiency while maintaining a certain fairness among the users.

The paper is organized as follows: Section 2 provides a literature review on fairness and utility function. Existing work on risk management in finance is presented in Section 3, and transposed to resource allocation in telecommunication networks in Section 4. Section 5 introduces the concept of unfairness aversion for the need of trade-off between fairness and efficiency when the limiting factor is the total rate. Simulations are presented in Section 6. Concluding remarks are given in Section 7.

2 Fairness and utility function

2.1 Fairness

Various approaches have been proposed for fairness in allocation of a single type of resource. Let $x = (x_1, ..., x_n)$ be a resource allocation vector. A variety of metrics, such as the ratio between the smallest and the largest allocations (min-max ratio), Jain's index, or proportional fairness have been proposed:

2.1.1 Min-max ratio

The min-max ratio is the ratio between the lowest allocation and the highest allocation.

$$\min - \max(x_1, ...x_n) = \frac{\min_i x_i}{\max_i x_i} \tag{1}$$

2.1.2 Jain's index

$$J(x_1, x_2, ..., x_n) = \frac{(\sum_{i=1}^n x_i)^2}{n \sum_{i=1}^n x_i^2} \tag{2}$$

In the worst case, when all x_i but one equal to 0, $J(x_1, x_2, ..., x_n) = \frac{1}{n}$. In the best case, when all x_i are equal, $J(x_1, x_2, ..., x_n) = 1$.

By noting $\hat{x}$ and $\hat{\sigma_x}^2$ the empirical mean and the empirical squared error of x, respectively, Jain's index can be expressed as a function of these two parameters:

$$J(x_1, x_2, ..., x_n) = \frac{1}{1 + \frac{\hat{\sigma_x}^2}{\hat{x}^2}} \tag{3}$$

2.1.3 Proportional fairness

Proportional fairness is based upon the assumption that the users' utility functions are logarithmic. Resource allocations x_i^* are proportionally fair if they maximize $\sum_{i=1}^n \log x_i$.

In weighted proportional fairness, the optimal allocation is obtained by maximizing $\sum_{i=1}^n p_i \log x_i$.

The concept of proportional fairness has been generalized in [21], which introduces the definition of (p, α)-proportional fairness: a feasible resource vector x^* is (p, α)-proportionally fair if for any feasible vector x, $\sum_{i=1}^n p_i \frac{x_i - x_i^*}{x_i^{*\alpha}} \leq 0$.

It is shown in [21] that (p, α)-proportional fair resource allocation is equivalent to the maximization of the function:

$$f_\alpha(x) = \begin{cases} \ln x, & \text{if } \alpha = 1 \\ \frac{x^{1-\alpha}}{1-\alpha}, & \text{otherwise} \end{cases} \tag{4}$$

This definition can be slightly modified in order to have a family of functions continuous in α:

$$f_\alpha(x) = \int_1^x \frac{dt}{t^\alpha} = \begin{cases} \ln x, & \text{if } \alpha = 1 \\ \frac{x^{1-\alpha}-1}{1-\alpha}, & \text{otherwise} \end{cases} \tag{5}$$

2.1.4 Entropy

Entropy was introduced by Shannon [27] in information theory in order to measure the expected value of the information contained in a message. Assuming that a random variable can take n distinct values, with probabilities p_1, p_2,...,p_n, Shannon entropy is defined as:

$$H(p_1, ..., p_n) = -\sum_{i=1}^n p_i \log_2 p_i \tag{6}$$

Shannon entropy equals to 0 when one probability equals to 1 and all other probabilities equal to 0. Shannon entropy reaches its maximum value, $\log_2 n$ when all probabilities equal to $\frac{1}{n}$. More generally, the more homogeneous the probability distribution is, the greater Shannon entropy is. For this reason, Shannon entropy can be used as a fairness measure. By defining:

$$p_i = \frac{x_i}{\sum_{j=1}^n x_j} \tag{7}$$

$H(p_1, ..., p_n)$ is a fairness measure of the resource allocation vector $x = (x_1, ..., x_n)$.

In order to unify these various theories, [13] and [9] developed an axiomatic approach which is summarized below. A fairness measure is a function

$f : \mathbb{R}^n \to \mathbb{R}$

where $x \in \mathbb{R}^n$ is an allocation vector representing the resource allocated to each user, fulfilling the following axioms:

- Continuity: f is continuous for any $n \geq 1$.
- Homogeneity: $f(x) = f(t.x), \forall t > 0$
- Saturation: As the number of users tends to infinity, fairness value of equal resource allocation becomes independent of the number of users: $\lim_{n\to\infty} \frac{1_n}{1_{n+1}}$.
- Partition:
  The partition axiom uses the concept of mean function:

  $h : \mathbb{R}^2 \to \mathbb{R}$

  is a mean function iff there exist a continuous and strictly monotonic function g and two positive weights s_1 and s_2 fulfilling $s_1 + s_2 = 1$ such that: $\forall u, v \in \mathbb{R}, g(h(u, v)) = s_1 g(u) + s_2 g(v)$.
  Considering an arbitrary partition of the system into two subsystems, the partition axiom states that there exists a mean function

  $h : \mathbb{R}^2 \to \mathbb{R}$

  such that the fairness ratio of two resource allocation vectors $x = [x^1 x^2]$ and $y = [y^1 y^2]$ equals the mean of the fairness ratios of the two suballocations: $\frac{f(x)}{f(y)} = h\left(\frac{f(x^1)}{f(y^1)}, \frac{f(x^2)}{f(y^2)}\right)$.
- Starvation: In a two-user system, an equal allocation is more fair than starving one user: $f([11]) \geq f([10])$.

Starting with these five axioms, [13] generates a family of fairness measures from a generator function:

$$f_\beta(x) = \mathrm{sign}(1-\beta)\left(\sum_{i=1}^{n}\left(\frac{x_i}{\sum_{j=1}^{n} x_j}\right)^{1-\beta}\right)^{\frac{1}{\beta}} \quad (8)$$

where sign(.) is the sign function.

Fairness in network resource allocation encounters the question of fair balance of utility functions, rather than fair balance of throughputs: fair balance of throughputs can lead to very unsatisfied users. For example, if two users have an all-or-nothing utility function for the total available throughput in the network, the best solution is to choose arbitrarily one of the users and to attribute him the whole throughput.

2.2 Utility function

On the other hand, maximization of utility function is also considered. These two approaches are clearly different, since scale-invariant metrics are unaffected by the total amount of allocated resources, while maximization of utility functions leads to Pareto optimal resource allocations.

In [6], users' satisfaction is defined taking into account the available resource. As a consequence, a user receiving a given allocation will have a greater satisfaction if all or part of the allocation he receives has been taken at the expense of other users. This definition leads to a trade-off between fairness and efficiency which favors compromise from the part of the users. This is fairly adapted to a resource sharing problem where users know all the allocations or at least the total allocation. However, the basic need of a user connecting to a telecommunication network is to transmit his data, without taking into account other users' allocations. Therefore, in our model, the utility function of a user only depends upon the throughput he receives.

We assume in the following that each user requests a given throughput R and has a utility function f fulfilling the following conditions:

- $f(0) = 0$
- $f(R) = 1$
- $\forall x \leq 0, f(x) = 0$
- $\forall x \geq R, f(x) = 1$
- $\forall x < R, f(x) < 1$
- f is a growing function

It has been shown in [29] that if the utility function is concave, then the general utility is maximized when resources are equally allocated. However, we will not make any assumption on convexity since all kinds of convexity properties can occur in practice for utility functions.

We assume that each user's utility function is known by the network, with no possibility of cheating.

Since the question is sharing one resource between several users, we will deal with cardinal utility rather than ordinal utility. The utility functions are considered according to their absolute values and treated as additive and multiplicative. Therefore, unlike most utility functions in use in economic modeling, the utility functions we consider are affected by composition with a growing function.

Many well-known resource allocation optimization policies are particular cases of utility functions. For example:

- a constraint requiring the $SINR$ to be greater than a threshold [14], [12] is equivalent to an all-or-nothing utility function;

– optimizing the total throughput [26] is equivalent to a linear utility function.
– proportional fairness [11], [20] is equivalent to a logarithmic utility function.

Taking into account the fact that each player tries to optimize his utility function, the fairness measure of a resource allocation vector $x = (x_1, ..., x_n)$ will be defined as Jain's index of utility functions:

$F(x_1, x_2, , x_n) = J\left(f_1(x_1), f_2(x_2), ..., f_n(x_n)\right)$

$F(x_1, x_2, , x_n) = \frac{\left(\sum_{i=1}^{n} f_i(x_i)\right)^2}{n \sum_{i=1}^{n} f_i(x_i)^2}$

In the case where there are several classes of users, Jain's index may be weighted:

$F(x_1, x_2, , x_n) = J\left(f_1(x_1), f_2(x_2), ..., f_n(x_n)\right)$

$F(x_1, x_2, , x_n) = \frac{\left(\sum_{i=1}^{n} \beta_i f_i(x_i)\right)^2}{n \sum_{i=1}^{n} (\beta_i f_i(x_i))^2}$

3 A PARALLEL WITH FINANCE

In the expected utility model, introduced by Daniel Bernoulli in the 18$^{\text{th}}$ century, a rational player behaves as if he tries to maximize his expected utility function. That is to say that if a rational player participates to a lottery A, where the gains are A_1,..., A_n with respective probabilities p_1,...,p_n, the player will attribute to this lottery a cardinal utility:

$\varphi(A) = \sum_{k=1}^{n} p_k f(A_k)$, where f is some growing function.

Then, a rational player is able to define his preference between two lotteries: if $\varphi(A) \geq \varphi(B)$, the player prefers lottery A to lottery B ($A \succeq B$).

This model was formally developed by John Von Neumann and Oscar Morgenstern [28], who stated four axioms which define a rational decision maker:

– Completeness: for every A and B, $A \succeq B$ or $B \succeq A$.
– Transitivity: for every A, B and C, if $A \succeq B$ and $B \succeq C$, then $A \succeq C$.
– Independence: for any $t \in [0, 1]$ and for every A, B and C, if $A \succeq B$, then $tA + (1-t)C \succeq tB + (1-t)C$.
– Continuity: for every A, B and C, if $A \succeq B \succeq C$, there exist some $t \in [0, 1]$ such that B is equally good as $tA + (1 - t)C$.

In 1953, Maurice Allais published simple examples disproving Bernoulli's utility function and Von Neumann and Morgenstern's independence assumption [1].

3.1 Utility function

Allais' paradox is based on the four lotteries described in Table 1[1].

[1] French francs were originally used in all the paradoxes

Table 1 Allais paradox

Lottery	Chance	Winnings
1A	100%	1 M$
1B	10%	5 M$
	89%	1 M$
	1%	0
2A	11%	1 M$
	89%	0
2B	10%	5 M$
	90%	0

Allais claims that a rational player can prefer lottery $1A$ to lottery $1B$ and lottery $2B$ to lottery $2A$. The reason is risk aversion: if the risk is low, the player will prefer the more secure choice. If the risk is high anyway, the player will try to maximize the risk premium. However, these preferences contradict the expected utility model:

If the player prefers Lottery $1A$ to Lottery $1B$, then:

$$f(1M\$) > 10\% f(5M\$) + 89\% f(1M\$) + 1\% f(0) \quad (9)$$

Therefore,

$$11\% f(1M\$) > 10\% f(5M\$) + 1\% f(0) \quad (10)$$

But if he prefers Lottery $2B$ to Lottery $2A$, then:

$$10\% f(5M\$) + 90\% f(0) > 11\% f(1M\$) + 89\% f(0) \quad (11)$$

So,

$$11\% f(1M\$) < 10\% f(5M\$) + 1\% f(0) \quad (12)$$

Inequalities (10) and (12) clearly contradict each other.

3.2 Independence assumption

The independence assumption is based on the following argument: the lottery $tA + (1 - t)C$ (resp. $tB + (1 - t)C$) can be performed in two steps: first we draw lots between lotteries A (resp. B) and C with probabilities t and $(1-t)$, and then we play the chosen lottery. If the probability t event occurs, then we play lottery A (resp. lottery B). If the probability $(1 - t)$ event occurs, then we play lottery C anyway. Since lottery A is assumed to be preferable to lottery B, the player shall prefer lottery $tA + (1 - t)C$ to lottery $tB + (1 - t)C$.

However, Allais points out that the choice must be *ex ante* and not *ex post*, and provides an example based upon the three lotteries described in Table 2.

Table 2 Independence assumption - Counter-example

Lottery	Chance	Winnings
A	100%	100 M$
B	98%	500 M$
	2%	0
C	100%	0

With $t = 1\%$, the combined lotteries are described in Table 3.

Table 3 Combined lotteries

Lottery	Chance	Winnings
tA+(1-t)C	1%	100 M$
	99%	0
tB+(1-t)C	0.98%	500 M$
	99.02%	0

Rational, but cautious, players may prefer lottery A to lottery B and lottery $tB + (1 - t)C$ to lottery $tA + (1 - t)C$. Once again, as in the first paradox, the reason is risk aversion.

Allais' main conclusion is that all the properties of the probability distribution must be taken into account for any rational choice involving risk.

3.3 Risk-return trade-off in portfolio management

The portfolio management model was developed by Harry Markowitz in 1952 [19]. Given a set of individual assets with their respective expected returns, volatilities and correlations, it can be proved that the set of feasible portfolios is bounded by a curve, which is called the minimum-variance frontier ([3], Chapter 7). The upper part of the minimum-variance frontier is called the efficient frontier of risky assets. Any rational investor shall define his risky portfolio on the efficient frontier (Fig. 1).

The choice of a risky portfolio on the efficient frontier is a matter of risk aversion. A risk-averse investor will choose a portfolio close to the global minimum-variance portfolio, in order to reduce risk. A non-risk-averse investor will choose a more risky portfolio, offering a higher expected return.

4 TRANSPOSITION TO RESOURCE ALLOCATION

Allais' paradoxes can be transposed to resource allocation in wireless networks.

4.1 Utility function

Let us consider the four following networks, proposing various rates for equally-weighted users (see Table 4).

Table 4 Allais paradox - Transposition to resource allocation

Network	Users	Rate
1A	100%	1 Mbit/s
1B	10%	5 Mbit/s
	89%	1 Mbit/s
	1%	0
2A	11%	1 Mbit/s
	89%	0
2B	10%	5 Mbit/s
	90%	0

While it is generally assumed that Network $1A$ is more efficient than Network $1B$, it is also assumed that Network $2B$ is more efficient than Network $2A$. However, as in the example above, these statements contradict the definition of network efficiency.

4.2 Independence assumption

Let us consider the three following networks, proposing various rates for equally-weighted users (see Table 5).

Table 5 Independence assumption - Transposition to resource allocation

Network	Users	Rate
A	100%	100 Mbit/s
B	98%	500 Mbit/s
	2%	0
C	100%	0

With $t = 1\%$, the resulting networks are described in Table 6.

Table 6 Combined networks

Network	Users	Rate
tA+(1-t)C	1%	100 Mbit/s
	99%	0
tB+(1-t)C	0.98%	500 Mbit/s
	99.02%	0

An operator may prefer Network A to Network B, and Network $tB + (1 - t)C$ to Network $tA + (1 - t)C$.

In finance, the most widely accepted explanation for Allais' paradox is risk aversion: if the risk is low, the player will prefer the more secure choice. If the risk

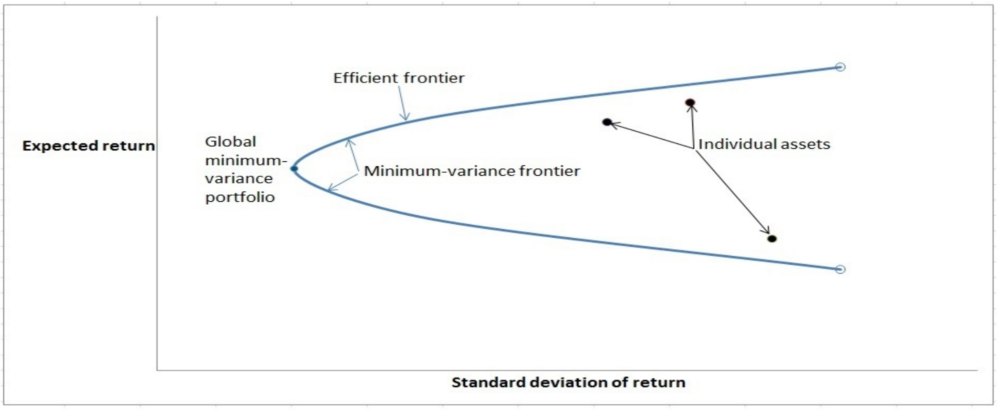

Fig. 1 The efficient frontier of risky assets.

is high anyway, the player will try to maximize the risk premium.

A similar explanation can be adopted for network efficiency: in Network $1B$, 10% fully satisfied users will be negatively overwhelmed by 1% of fully unsatisfied users. Network $1A$ will be preferred because it does not let any user fully unsatisfied.

On the other hand, the difference of fully unsatisfied users between Networks $2A$ and $2B$ is not significant (89% vs 90%). For 1% more unsatisfied users, it seems worthy to enhance the satisfaction level of other users.

Therefore, as in the example of utility function in finance, network efficiency shall be considered as a prescriptive theory, rather than a descriptive theory.

A major consequence on this paradox is that resource allocation in a network must be made by considering the network as a whole. Sharing the network into subnetworks and optimizing resource allocation in each one of them will lead to suboptimal allocation.

5 FAIRNESS-EFFICIENCY TRADE-OFF

A survey of fair optimization methods and models for resource allocation in telecommunication networks is provided in [23]. The concept of trade-off between fairness and utility is mentioned in [8]. A metric to evaluate the price of fairness in terms of efficiency loss and the price of efficiency in terms of fairness loss is provided in [2]. A method of management of the efficiency-fairness trade-off by controlling the system fairness index is proposed in [25]. An α-fair dynamically adapted scheduling strategy optimizing coverage and capacity in self-organizing networks is proposed in [4]. In order to introduce our own definition of fairness-efficiency trade-off, we will first define the concepts of network efficiency and unfairness aversion.

5.1 Network efficiency

A simple measure for network efficiency can be the total binary rate of the network. However, we prefer to relate network efficiency to the user's utility functions. The reason is that a network can be inefficient even though the binary rate may be high. For example, if two users have an all-or-nothing utility function for the total available throughput in the network, choosing arbitrarily one of the users and attributing him the whole throughput is a more efficient solution than sharing the throughput into two equal parts.

Optimizing the total binary rate is a particular case of our model, since it is equivalent to rule that the users' utility functions are linear.

Therefore, we define the network efficiency as the sum, optionally weighted, of the user's utility functions:

$$U(x_1, x_2, ..., x_n) = \sum_{i=1}^{n} \alpha_i f_i(x_i) \tag{13}$$

where x_i is the rate allocated to user i and f_i his utility function.

5.2 Unfairness aversion

The definition we adopt for unfairness aversion is similar to that of risk aversion in finance.

Unfairness aversion is the reluctance of a network to accept a given unfairness level for a given efficiency level rather than a lower unfairness level for a lower efficiency level.

The concept of unfairness aversion gives a measure of the price the network is ready to pay in terms of efficiency in order to get a better fairness.

Therefore, the network's indifference curves are the set of points in the efficiency-fairness map which are equally satisfying from the network's point of view.

Thus, the network's characteristics can be represented by a point in a two-axis map. Among all the feasible points, the fairness-efficiency trade-off is the point located on the highest indifference curve.

If the fairness is defined by Jain's index, which can be, as mentioned above, expressed as a function of empirical mean and empirical squared error, then the fairness-efficiency trade-off is equivalent to the risk-return trade-off in finance. Though mean and standard deviation do not provide all the information about probability distribution, these two parameters may provide a fairly good approximation in order to describe the operator's preferences.

Unfairness aversion can be characterized by an objective function $\varphi(U, J)$ where the curves $\varphi(U, J) = C$ are the indifference curves, C denoting a constant.

5.3 Trade-off between fairness and efficiency

The optimal trade-off between fairness and efficiency is given by the constrained optimization:

$$\max \varphi\left(U(x_1, x_2, ..., x_n), J(x_1, x_2, ..., x_n)\right) \quad (14)$$

subject to $\sum_{i=1}^{n} x_i = X$, X being the total allocated resource.

The approach described above can be illustrated by the following examples. In these examples, two kinds of utility functions will be used:

- linear utility function: the user's satisfaction is proportional to the rate he gets;
- all-or-nothing utility function: the user is satisfied if and only if he gets the rate he wants.

In practice, linear utility functions can be encountered for example for data transmission, where the time transmission will be inversely proportional to the rate, while all-or-nothing utility functions can be met for real-time video applications.

6 Simulations

We ran simulations of the approach presented above in a network involving a large number of users.

6.1 System description

Users arrive into the network according to a Poisson process. Each user i is characterized by 4 parameters:

- Utility function (linear or all-or-nothing)
- Requested rate: R_i for an all-or-nothing user, R'_i for a linear user. The requested rate follows a lognormal law
- Volume of data which needs to be transmitted: Q_i for an all-or-nothing user, Q'_i for a linear user. The volume of data follows a lognormal law.
- User class (gold, silver, bronze)

We denote r_i (resp. r'_i) the rate allocated to all-or-nothing (resp. linear) user i and q_i (resp. q'_i) the remaining data to be transmitted by all-or-nothing (resp. linear) user i. At each instant, the number of all-or-nothing users is n and the number of linear users is m. The total number of users is $N = n + m$.

An event is the entry of a newcomer or the exit of a user when his transmission is achieved.

The system will process the resource allocation which optimizes the fairness-efficiency trade-off, under the constraint that the total rate is R_{total}.

The system will provide statistical outputs regarding the rate of satisfaction according to the type of users (class, demand or utility function).

- If the utility function of user i is all-or nothing, it is defined as:
  - $f_i(r_i) = 1_{[R_i, +\infty[}$
- If the utility function of user i is linear, it is defined as:
  - $f'_i(r'_i) = r'_i / R'_i$ for $0 \leq x_i \leq R'_i$
  - $f'_i(r'_i) = 1$ for $r'_i \geq R'_i$

In order to define the priorities among the user classes, utility functions are weighted for the calculation of the network efficiency and Jain's index.

The network efficiency is:

$U(r_1, , r_n, r'_1, ...r'_m) = \sum_{i=1}^{n} \alpha_i f_i(r_i) + \sum_{i=1}^{m} \alpha'_i f'_i(r'_i)$

where:

- $\alpha_i = 9$ or $\alpha'_i = 9$ if user i is a gold user;
- $\alpha_i = 3$ or $\alpha'_i = 3$ if user i is a silver user;
- $\alpha_i = 1$ or $\alpha'_i = 1$ if user i is a bronze user.

Jain's index is:

$F(r_1, , r_n, r'_1, ...r'_m) = J\left(f_1(r_1), ..., f_n(r_n), f'_1(r'_1), ..., f'_m(r'_m)\right)$

$F(r_1, , r_n, r'_1, ...r'_m) = \frac{\left(\sum_{i=1}^{n} \beta_i f_i(r_i) + \sum_{i=1}^{m} \beta'_i f'_i(r'_i)\right)^2}{N\left(\sum_{i=1}^{n} (\beta_i f_i(r_i))^2 + \sum_{i=1}^{m} (\beta'_i f'_i(r'_i))^2\right)}$,

where:

- $\beta_i = 1$ or $\beta'_i = 1$ if user i is a gold user
- $\beta_i = 3$ or $\beta'_i = 3$ if user i is a silver user

- $\beta_i = 9$ or $\beta'_i = 9$ if user i is a bronze user

We deal with three kinds of indifference curves:

- $\phi(U, J) = U(r_1, , r_n, r'_1, ...r'_m)$;
- $\phi(U, J) =$ $U(r_1, , r_n, r'_1, ...r'_m) J\left(f_1(r_1), ..., f_n(r_n), f'_1(r'_1), ..., f'_m(r'_m)\right)^\eta$, where η is a strictly positive number;
- $\phi(U, J) = J\left(f_1(r_1), ..., f_n(r_n), f'_1(r'_1), ..., f'_m(r'_m)\right)$.

For each event, the system will process the optimal allocation:

$$\max_{\sum_i r_i \leq R_{total}} \phi(U, J)$$

6.2 Algorithm

Any all-or-nothing user shall be either fully served or not served at all. As a result, all-or-nothing users bring discontinuities in the network efficiency and Jain's index, making impossible the use of classical algorithms for optimization.

Taking into account all the possibilities to serve or not serve each all-or-nothing user would lead to an exponentially complex problem.

Therefore, we present hereafter an approximation which enables to find a sub-optimal solution with a polynomial complexity. Under reasonable assumptions, this sub-optimal solution will tend to the optimal solution when the number of users tends to infinity.

6.2.1 Ordering the all-or-nothing users

The all-or-nothing users will be ordered according to the $\frac{\alpha_k}{R_k}$ ratio. The purpose of comparing these ratios is to evaluate the average impact of each unit of rate on the network efficiency.

After this ordering, we have n ordered all-or-nothing users and m linear users, with $n + m = N$.

$$X_1 = \begin{cases} R_1 \\ Q_1 \\ \alpha_1 \end{cases}, X_2 = \begin{cases} R_2 \\ Q_2 \\ \alpha_2 \end{cases}, \ldots, X_n = \begin{cases} R_n \\ Q_n \\ \alpha_n \end{cases},$$

with $\frac{\alpha_1}{R_1} \geq \frac{\alpha_2}{R_2} \geq ... \geq \frac{\alpha_n}{R_n}$

$$X_{n+1} = \begin{cases} R_{n+1} \\ Q_{n+1} \\ \alpha_{n+1} \end{cases}, X_{n+2} = \begin{cases} R_{n+2} \\ Q_{n+2} \\ \alpha_{n+2} \end{cases}, \ldots, X_{n+m} = \begin{cases} R_{n+m} \\ Q_{n+m} \\ \alpha_{n+m} \end{cases}$$

6.2.2 Basic assumption: priority among all-or-nothing users

We now assume that if an all-or-nothing user X_k, $1 \leq k \leq n$ is served, then all the all-or-nothing users who are better rated than him (ie X_i, $1 \leq i \leq k$) are also served. This assumption is an approximation. Since rate demands are not breakable, it may lead to a sub-optimal solution. The priority that we define among all-or-nothing users is based on the following considerations:

- **Lemma 1** *Let X_j and X_k be two all-or-nothing users. If $\alpha_j = \alpha_k$ and $R_j \leq R_k$, then:*

$$\max_{\substack{\sum_i r_i \leq R_{total} \\ r_j = R_j \\ r_k = 0}} \phi(U, J) \geq \max_{\substack{\sum_i r_i \leq R_{total} \\ r_j = 0 \\ r_k = R_k}} \phi(U, J)$$

Proof

$$\max_{\substack{\sum_i r_i \leq R_{total} \\ r_j = R_j \\ r_k = 0}} \phi(U, J) = \max_{\substack{\sum_{i \notin \{j,k\}} r_i \leq R_{total} - R_j \\ r_j = R_j \\ r_k = 0}} \phi(U, J)$$

$$\max_{\substack{\sum_i r_i \leq R_{total} \\ r_j = 0 \\ r_k = R_k}} \phi(U, J) = \max_{\substack{\sum_{i \notin \{j,k\}} r_i \leq R_{total} - R_k \\ r_j = 0 \\ r_k = R_k}} \phi(U, J)$$

Since $U(r_1, ..., r_j = R_j, ..., r_k = 0, ..., r_n, r'_1, ...r'_m) = U(r_1, ..., r_j = 0, ..., r_k = R_k, ..., r_n, r'_1, ...r'_m)$ and $F(r_1, ..., r_j = R_j, ..., r_k = 0, ..., r_n, r'_1, ...r'_m) = F(r_1, ..., r_j = 0, ..., r_k = R_k, ..., r_n, r'_1, ...r'_m)$, the lemma results from the fact that $R_{total} - R_j \geq R_{total} - R_k$.

- **Lemma 2** *Let X_{j_1},..., X_{j_q} be q all-or-nothing users with $\alpha_{j_1} = ... = \alpha_{j_q} = \alpha_j$ and X_{k_1},..., X_{k_p} p all-or-nothing users with $\alpha_{k_1} = ... = \alpha_{k_p} = \alpha_k$. If $\frac{\alpha_j}{\alpha_k} = \frac{p}{q}$, then $U(r_1, ..., r_{j_1} = R_{j_1}, ..., r_{j_q} = R_{j_q}, r_{k_1} = 0, ..., , r_{k_p} = 0, ..., r_n, r'_1, ...r'_m) = U(r_1, ..., r_{j_1} = 0, ..., r_{j_q} = 0, r_{k_1} = R_{k_1}, ..., , r_{k_p} = R_{k_p}, ..., r_n, r'_1, ...r'_m)$*

Proof $U(r_1, ..., r_{j_1} = R_{j_1}, ..., r_{j_q} = R_{j_q}, r_{k_1} = 0, ..., , r_{k_p} = 0, ..., r_n, r'_1, ...r'_m) = q\alpha_j + \sum_{i = \notin \{j_1, ...j_q, k_1, ...k_p\}} \alpha_i f_i(r_i) + \sum_{i=1}^m \alpha'_i f'_i(r'_i)$
$U(r_1, ..., r_{j_1} = 0, ..., r_{j_q} = 0, r_{k_1} = R_{k_1}, ..., , r_{k_p} = R_{k_p}, ..., r_n, r'_1, ...r'_m) = p\alpha_k + \sum_{i = \notin \{j_1, ...j_q, k_1, ...k_p\}} \alpha_i f_i(r_i) + \sum_{i=1}^m \alpha'_i f'_i(r'_i)$
Since $q\alpha_j = p\alpha_k$, the equality is obtained.

Lemma 2 shows that there are two groups of users which are equivalent with regard to the network efficiency. Of course, these two groups may have a different impact on Jain's index. However, the expression of Jain's index shows that if the number of other users is much larger than the size of these two groups, serving either one of these two groups should have little effect on Jain's index. This gives legitimacy to ordering the all-or-nothing users according to the $\frac{\alpha_k}{R_k}$ ratio, since this ratio represents the gain in network efficiency relative to the allocated rate.

This approximation enables to reduce the number of optimizations from 2^n to $n+1$.

6.2.3 The algorithm

The simulation algorithm is described in Figure 2.

6.2.4 Simulation parameters

The simulations were run with the following parameters:

- $R_{total} = 1,000$ Mbit/s
- Number of events: 4,000.
- Requested rate $R_i = R_0 2^{u_i}$ or $R'_i = R_0 2^{u'_i}$, with $R_0 = 100$ Mbit/s , and u_i and u'_i following a standard normal distribution.
- Volume of data which needs to be transmitted: $Q_i = Q_0 2^{v_i}$ or $Q'_i = Q_0 2^{v'_i}$, with $Q_0 = 1,000$ Mbit, and v_i and v'_i following a standard normal distribution.
- Newcomers arrive according to a Poisson process. The Poisson parameter λ is calculated in order to have a stable long-term number of users in the network.

  If n users, each one willing to transmit a volume of data $Q_i = Q_0 2^{v_i}$, $1 \leq i \leq n$ are in the network, the average requested time to exit all of them is:

$$T = \overline{\sum_{i=1}^{n} \frac{Q_i}{R_{total}}} = \frac{Q_0}{R_{total}} \overline{\sum_{i=1}^{n} 2^{v_i}} = \frac{Q_0}{R_{total}} \overline{\sum_{i=1}^{n} e^{v_i \ln 2}}$$

$$T = \frac{nQ_0}{R_{total}} exp(\ln 2^2/2) \quad (15)$$

  Therefore, in order that the average number of incoming users equals the average number of exiting users, $\lambda = \frac{R_{total}}{Q_0 exp(\ln 2^2/2)} \approx 0.78645$.
- All random processes are independent.

Fairness-efficiency trade-off : In order to evaluate the impact of the fairness-efficiency trade-off on the network performance, we ran the simulations with three different objective functions:

- $\phi(U,J) = U$; maximization of network efficiency;
- $\phi(U,J) = UJ$; trade-off between fairness and efficiency, with $\eta = 1$;
- $\phi(U,J) = J$; maximization of fairness.

It should be noted that the resource which is allocated is not storable. For this reason, the efficiency and the fairness are measured at any time for the users who are connected to the network at this time. There is no fairness between users connected at different times to the network.

6.2.5 Results

The minimum time a user i stays in the network is $\frac{Q_i}{R_i}$ or $\frac{Q'_i}{R'_i}$. Therefore, a relevant parameter to evaluate the network performance, from the point of view of user i, is $x_i = \frac{Q_i}{t_i R_i}$ or $x'_i = \frac{Q'_i}{t'_i R'_i}$, where t_i (resp. t'_i) is the delay between the arrival of the all-or-nothing user (resp. linear user) i to the network and the end of his transaction.

$\phi(\mathbf{U},\mathbf{J}) = \mathbf{U}$

Figures 3, 4 and 5 provide the cumulative distribution functions of x_i resulting from the simulations for each category of users. Means and standard deviations are summarized in Table 7.

For each class of users, cumulative distribution functions, means and standard deviation are similar for all-or-nothing utility function users and linear utility function users.

AON Users	Mean	Standard deviation
AON Gold	0.98	0.14
AON Silver	0.89	0.25
AON Bronze	0.57	0.39
LinearGold	0.97	0.12
Linear Silver	0.87	0.26
Linear Bronze	0.55	0.37

Table 7 Simulations results $\phi(U,J) = U$.

$\phi(\mathbf{U},\mathbf{J}) = \mathbf{UJ}$

Figures 6, 7 and 8 provide the cumulative distribution functions of x_i resulting from the simulations for each category of users. Means and standard deviations are summarized in Table 8.

For each class of users, the number of fully-satisfied users ($x_i = 1$) is much greater for all-or-nothing utility function users. The standard deviations and the numbers of presumably disappointed customers (low values of x_i) are much greater for all-or-nothing utility function users.

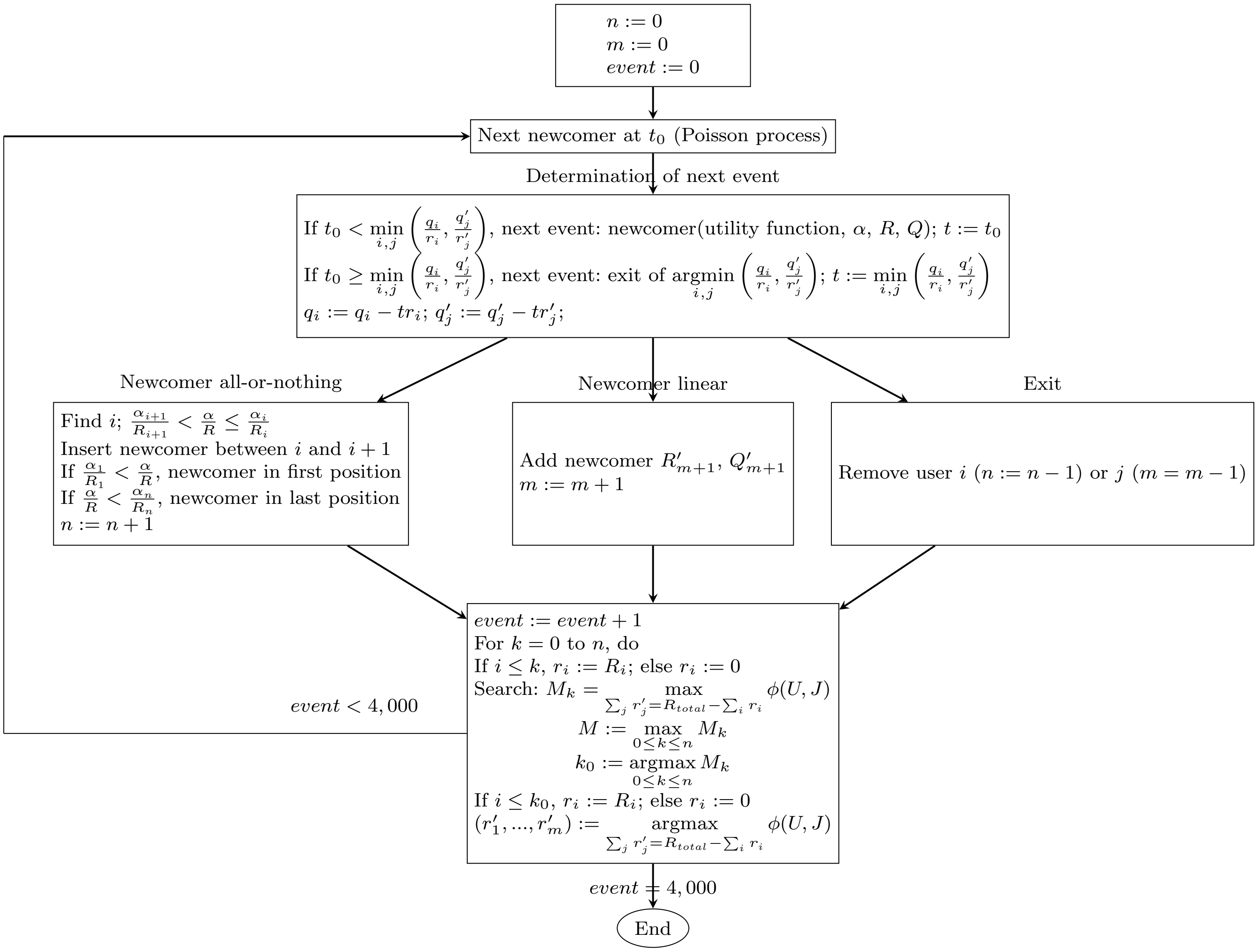

Fig. 2 Simulation Algorithm.

Users	Mean	Standard deviation
AON Gold	0.93	0.21
AON Silver	0.78	0.32
AON Bronze	0.39	0.30
Linear Gold	0.94	0.14
Linear Silver	0.76	0.24
Linear Bronze	0.39	0.15

Table 8 Simulations results $\phi(U, J) = UJ$.

$\phi(\mathbf{U}, \mathbf{J}) = \mathbf{J}$

Figures 9, 10 and 11 provide the cumulative distribution functions of x_i resulting from the simulations for each category of users. Means and standard deviations are summarized in Table 9.

For each class of users, the number of fully-satisfied users ($x_i = 1$) is much greater for all-or-nothing utility function users. The standard deviations and the numbers of presumably disappointed customers (low values of x_i) are much greater for all-or-nothing utility function users.

Users	Mean	Standard deviation
AON Gold	0.67	0.36
AON Silver	0.35	0.32
AON Bronze	0.17	0.14
Linear Gold	0.55	0.28
Linear Silver	0.34	0.15
Linear Bronze	0.15	0.06

Table 9 Simulations results $\phi(U, J) = J$.

As the comparison between Tables 7, 8 and 9 shows, the performance is strongly degraded when the network maximizes the fairness ($\phi(U, J) = J$). When comparing the performances between maximizing the network efficiency ($\phi(U, J) = U$) and maximizing a trade-

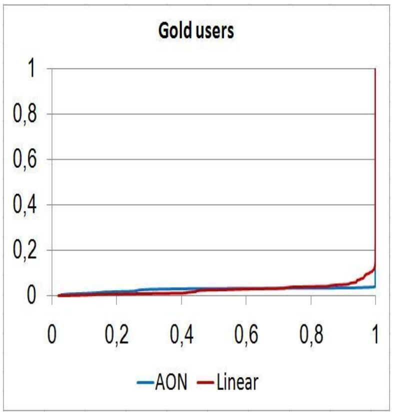

Fig. 3 Gold users.

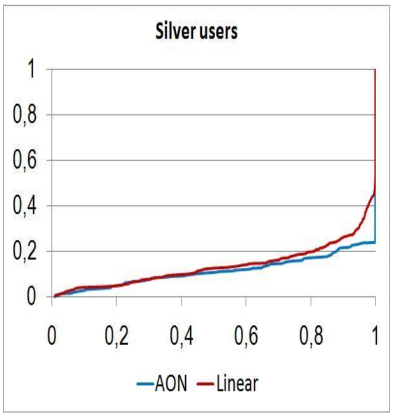

Fig. 4 Silver users.

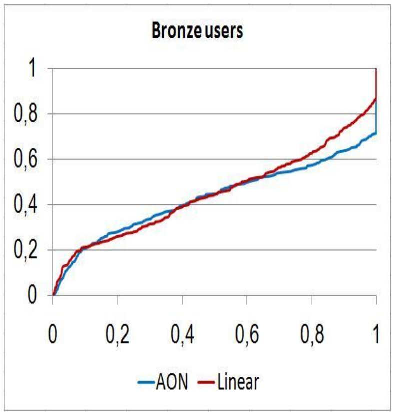

Fig. 5 Bronze users.

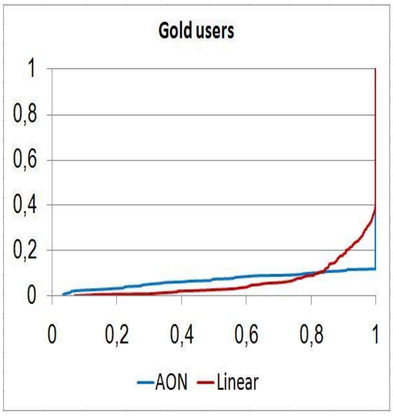

Fig. 6 Gold users.

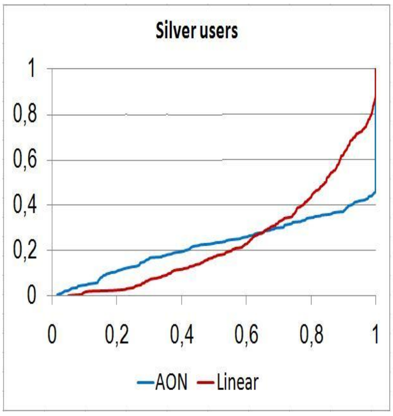

Fig. 7 Silver users.

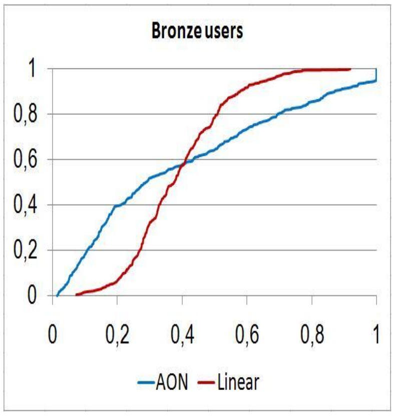

Fig. 8 Bronze users.

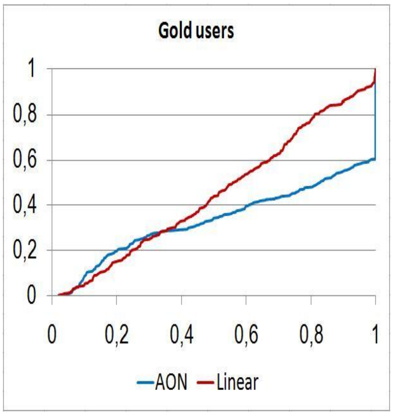

Fig. 9 Gold users.

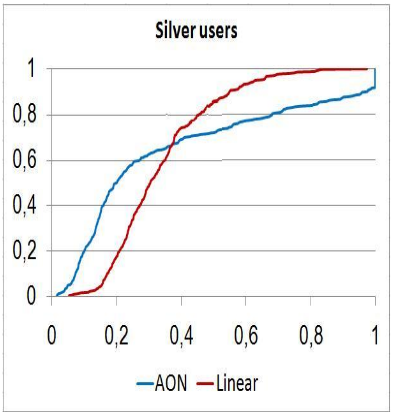

Fig. 10 Silver users.

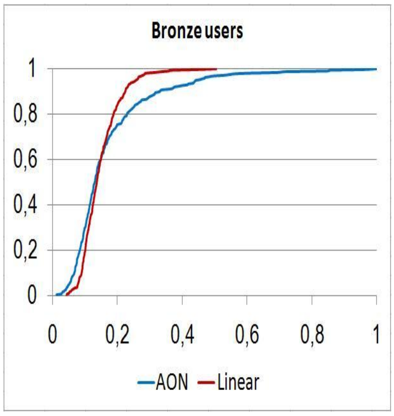

Fig. 11 Bronze users.

off between fairness and efficiency ($\phi(U, J) = UJ$), it turns out that the means of x_i are always higher for $\phi(U, J) = U$.

However, as the comparison between the cumulative distribution functions of x_i shows, the number of presumably disappointed customers (low values of x_i) is also significantly greater.

This example highlights the fact that, by defining its fairness-efficiency trade-off function and adjusting the parameter η in the objective function $\phi(U, J) = UJ^\eta$, an operator can define its priorities and find the proper balance between maximizing the network efficiency and minimizing the number of disappointed customers.

7 Conclusion

Tools already existing in finance can be successfully transposed to resource allocation in telecommunication networks. Drawing a parallel with the Bernoulli model, which is disproved by Allais' paradox, we challenge the common approach based on utility function maximization for resource allocation. By introducing the concepts of network efficiency and unfairness aversion, we propose a model based on fairness-efficiency trade-off, similar to the well-known risk-return trade-off in finance. This approach enables to define an optimal resource allocation without conflicting with paradoxes resulting from the maximization of a utility function.

Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Allais, M.: The so-called Allais paradox and rational decisions under uncertainty. In: Expected utility hypotheses and the Allais paradox, pp. 437–681. Springer (1979)
2. Bertsimas, D., Farias, V.F., Trichakis, N.: On the efficiency-fairness trade-off. Management Science **58**(12), 2234–2250 (2012)
3. Bodie, Z., Kane, A., Marcus, A.J.: Investment and portfolio management. McGraw-Hill Irwin (2011)
4. Combes, R., Altman, Z., Altman, E.: On the use of packet scheduling in self-optimization processes: application to coverage-capacity optimization. In: Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), 2010 Proceedings of the 8th International Symposium on, pp. 98–107. IEEE (2010)
5. Feng, D., Jiang, C., Lim, G., Cimini, L.J., Feng, G., Li, G.Y.: A survey of energy-efficient wireless communications. IEEE Communications Surveys & Tutorials **15**(1), 167–178 (2013)
6. Fossati, F., Moretti, S., Secci, S.: A mood value for fair resource allocations. arXiv preprint arXiv:1701.07991 (2017)
7. Hong, M., Garcia, A.: Mechanism design for base station association and resource allocation in downlink OFDMA network. IEEE Journal on Selected Areas in Communications **30**(11), 2238–2250 (2012)
8. Huaizhou, S., Prasad, R.V., Onur, E., Niemegeers, I.: Fairness in wireless networks: Issues, measures and challenges. IEEE Communications Surveys & Tutorials **16**(1), 5–24 (2014)
9. Joe-Wong, C., Sen, S., Lan, T., Chiang, M.: Multiresource allocation: Fairness-efficiency tradeoffs in a unifying framework. IEEE/ACM Transactions on Networking (TON) **21**(6), 1785–1798 (2013)
10. Kang, X., Zhang, R., Motani, M.: Price-based resource allocation for spectrum-sharing femtocell networks: A stackelberg game approach. IEEE Journal on Selected areas in Communications **30**(3), 538–549 (2012)
11. Kelly, F.: Charging and rate control for elastic traffic. European transactions on Telecommunications **8**(1), 33–37 (1997)
12. Koutsopoulos, I., Tassiulas, L.: Adaptive resource allocation in SDMA-based wireless broadband networks with OFDM signaling. In: INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, vol. 3, pp. 1376–1385. IEEE (2002)
13. Lan, T., Kao, D., Chiang, M., Sabharwal, A.: An axiomatic theory of fairness in network resource allocation. IEEE (2010)
14. Le, L., Hossain, E.: Multihop cellular networks: Potential gains, research challenges, and a resource allocation framework. IEEE Communications Magazine **45**(9), 66–73 (2007)
15. Lee, J.W., Mazumdar, R.R., Shroff, N.B.: Joint resource allocation and base-station assignment for the downlink in CDMA networks. IEEE/ACM Transactions on Networking **14**(1), 1–14 (2006)
16. Li, C., Song, S., Zhang, J., Letaief, K.B.: Maximizing energy efficiency in wireless networks with a minimum average throughput requirement. In: 2012 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1130–1134. IEEE (2012)
17. Lin, Y., Bao, W., Yu, W., Liang, B.: Optimizing user association and spectrum allocation in HetNets: A utility perspective. IEEE Journal on Selected Areas in Communications **33**(6), 1025–1039 (2015)
18. Maric, I., Bostjancic, B., Goldsmith, A.: Resource allocation for constrained backhaul in picocell networks. In: Information Theory and Applications Workshop (ITA), 2011, pp. 1–6. IEEE (2011)
19. Markowitz, H.: Portfolio selection. The journal of finance **7**(1), 77–91 (1952)
20. Massoulié, L., Roberts, J.: Bandwidth sharing: objectives and algorithms. In: INFOCOM'99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, vol. 3, pp. 1395–1403. IEEE (1999)
21. Mo, J., Walrand, J.: Fair end-to-end window-based congestion control. IEEE/ACM Transactions on Networking (ToN) **8**(5), 556–567 (2000)
22. Ng, D.W.K., Lo, E.S., Schober, R.: Energy-efficient resource allocation in OFDMA systems with hybrid energy harvesting base station. IEEE Transactions on Wireless Communications **12**(7), 3412–3427 (2013)
23. Ogryczak, W., Luss, H., Pióro, M., Nace, D., Tomaszewski, A.: Fair optimization and networks: A survey. Journal of Applied Mathematics **2014** (2014)
24. Pujol, F.: Mobile traffic forecasts 2010-2020 & offloading solutions. IDATE Consulting and Research, May **15** (2011)
25. Rodrigues, E.B., Casadevall, F.: Control of the trade-off between resource efficiency and user fairness in wireless networks using utility-based adaptive resource allocation. IEEE Communications Magazine **49**(9) (2011)
26. Seong, K., Mohseni, M., Cioffi, J.M.: Optimal resource allocation for OFDMA downlink systems. In: 2006 IEEE International Symposium on Information Theory, pp. 1394–1398. IEEE (2006)
27. Shannon, C.E.: A mathematical theory of communication. ACM SIGMOBILE Mobile Computing and Communications Review **5**(1), 3–55 (2001)
28. Von Neumann, J., Morgenstern, O.: Theory of games and economic behavior. Princeton university press (2007)
29. Ye, Q., Rong, B., Chen, Y., Al-Shalash, M., Caramanis, C., Andrews, J.G.: User association for load balancing in heterogeneous cellular networks. IEEE Transactions on Wireless Communications **12**(6), 2706–2716 (2013)
30. Yetgin, H., Cheung, K.T.K., El-Hajjar, M., Hanzo, L.H.: A survey of network lifetime maximization techniques in wireless sensor networks. IEEE Communications Surveys & Tutorials **19**(2), 828–854 (2017)