This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2022.3231840

Sharing data plans for cellular mobile data access

M. Karaliopoulos[1], G. Cheirmpos[2] and I. Koutsopoulos[1]

[1]Department of Informatics, Athens University of Economics and Business, Athens, Greece
[2]Research Content Operations team, Elsevier, Amsterdam, The Netherlands

***Abstract*—The demand for mobile data has been steadily increasing over the last decade, forming an ever-increasing portion of the overall Internet traffic. A great portion of this demand is still served through capped cellular data plans that charge a fixed fee for data consumption up to a cap and impose a typically higher penalty rate for consumption beyond that cap. It has been shown that when capped plans are shared, their caps are better utilized and the incurred penalty costs are amortized. This translates to subscription cost savings for the mobile users and better use of the cellular network resources for the mobile network operators. However, this sharing is nowadays restricted to *closed* groups (*e.g.*, family members) or to multiple devices of a single user.**

In this paper we explore the generalization of capped data plan sharing to *open* user groups. We take the viewpoint of a platform that seeks to organize cellular users into subscription groups and *recommend* to them shared data plans on offer by mobile network operators that maximize their subscription cost savings. We first introduce a new cost-sharing rule, called double proportional cost sharing (DPCS), for splitting the subscription charges of the shared capped data plans "fairly" between subscription group members. We then formulate the two platform tasks into a joint optimization problem, characterize its complexity and devise three algorithms that leverage clustering techniques to solve it. Under ideal prediction of users' data consumption all three algorithms achieve subscription savings beyond 50% for at least 70% of users and smaller but still significant savings for the rest of them, which are independent of the number of subscribers in all scenarios of practical interest. Notably, the best of the three algorithms preserves those savings when there is up to 10% bias in predicting the users' data consumption and when this consumption exhibits elasticity to the data cap.

***Index Terms*—Shared cellular data plans, mobile Internet access, clustering, pricing, optimization.**

I. Introduction

Reports by Internet commercial actors [1] and independent regulators [2] highlight similar persistent trends about mobile data consumption. On the one hand, mobile data has become the main vehicle for voice and text services substituting traditional voice and SMS services that steadily drop by millions minutes/messages per year. On the other hand, mobile video accounts for an ever increasing portion of mobile data. As a result, the total mobile data traffic volume grows, maybe faster in non-saturated markets, and the share of mobile data charges steadily increases in the operators' revenue breakdown.

In this context, the engineering of cellular data plans is viewed as a valuable tool for shaping the demand for mobile data and it facilitates the long-term network planning and resource management. Although unlimited data plans are now on offer by almost every cellular network operator, a great part of the mobile cellular data traffic is still realized through *capped* data plans [3]. Capped plans charge a fixed fee for consuming up to a predefined volume of data (*cap*) and a typically higher penalty fee (*overage charges*) for data volumes that exceed this cap. The stochasticity in the user data consumption patterns together with the relatively coarse granularity of caps in the offered cellular data plans generate inefficiencies in their actual usage: many of those caps are underutilized, whereas others are exceeded giving rise to overage charges. For the end users, this means unnecessary costs that, in principle, could have been avoided. The operator, on the other hand, may see a short-term benefit from collected overage charges but it needs to account for customer dissatisfaction and higher uncertainty in network resource planning.

The relevant literature has highlighted these inefficiencies and has early identified the sharing of data plans as a promising way to mitigate them [4] [5] [6]. To sum it up, subscribing to shared data plans, cellular users can collectively make better use of typically "bigger" data plans (*i.e.*, with higher cap) than those they individually subscribe to and save on their subscription costs. At the same time, they can assist the operator in network planning and may end up growing its subscriber base in the long run, not least because they could be perceived as evidence of flexible and innovative user-centered thinking. Nevertheless, shared data plans have been so far mainly considered in the context of closed groups (*e.g.*, family members) or individual subscribers owning multiple mobile devices. In this paper, we explore the generalization of these plans to more *open* groups of users such as the subscribers of a platform that issues recommendations about such plans.

The key challenges to this end relate to the actual partitioning of users into subscription-sharing groups, the way the group members will share the fixed and any overage charges resulting from their data consumption, and the ultimate assignment of the subscription-sharing groups to cost-effective shared capped data plans. Partitioning problems are combinatorial in nature, hence the ultimate goal of our work is to devise efficient well-performing algorithms for this *joint user partitioning and data plan assignment problem*. Such algorithms could lie at the core of third-party online platforms, possibly owned and maintained *e.g.*, by non-profit regional/national consumer associations, which issue recommendations to cellular subscribers for data plan sharing opportunities that save on their subscription costs.

A major part of this work was carried out while G. Cheirmpos was with the Department of Informatics, Athens University of Economics and Business. Contact emails: {mkaralio, jordan}@aueb.gr.

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2022.3231840

A. Related work

Various aspects of (cellular) data plan design have been subject to research over the last decade. In [7], Chen and Huang compare time-, volume- and rate-based pricing. Assuming a monopoly setting and users with heterogeneous utilities, they show that pricing users by Mbps facilitates congestion control and maximizes profits. In [8], Zheng *et al.* construct a quite elaborate model of how users adapt their daily data consumption to their residual data quota. This model helps identify which data plans are most beneficial for different types of users and, ultimately, dictates the offer of data plans by the ISP. Optimal data plan caps and subscription fees are also pursued in [9] with a contract-theoretic approach, this time for data plan structures (rollover and credit data plans) that provide end users with time flexibility. Wang *et al.* infer what subscribers are willing to pay and allocate them to plans with larger data caps that simultaneously maximize their utility and the operator profits. Rollover plans and their optimal offer timing are the focus point also in [10], in a competitive setting with multiple operators. It is shown that, depending on its market share, an MNO should take a leading or follower's role in announcing new plans, to better balance the immediate suffered revenue loss with the attraction of new users. Finally, in a more theoretical work with bolder assumptions [11], the authors propose a super flexible scheme with prediction of users' demand on a daily basis and assignment of a different plan each day by a Mobile Virtual Network Operator.

More relevant to our work are references [4], [5], and [6] that address shared data plans, in particular. In the first two cases, the emphasis is more on sharing across user's devices. Sen *et al.* in [4] analyze the choice between shared and individual capped data plans. Under simple assumptions about the users' consumption they show that the optimal choice depends heavily on how much end users reduce their data consumption when they exceed the cap. On the contrary, Jin and Pang [5] work with unlimited data plans drawing on the bundling model in [12]. They derive threshold conditions about the unit cost of service that render sharing profitable, both under independent and complementary user utilities on the different devices. Finally, Cardona *et al.* in [6] lie closer to the work in this paper focusing on the sharing of capped data plans between multiple users. Besides unfolding motivation for shared capped data plans, their study is the single one we are aware of that identifies the task of grouping users into subscription groups. However, they only postulate that good groupings comprise users who feature similar average values of data consumption without elaborating further into the algorithmic challenges of the grouping task. They rather focus on a different model, where individual users may sell/buy spare/extra capacity to/from the operator at a secondary market price. They argue that such a model lets operators with some control over data sharing when compared to the fully uncontrolled tethering of cellular data connections.

B. Our contributions

Three are the main contributions of our work:

First, we introduce a new cost-sharing rule, called double proportional cost sharing (DPCS), for splitting the subscription charges of the shared capped data plans "fairly" between subscription group members. Contrary to existing general-purpose cost-sharing rules, DPCS is tailored to the capped data plan particularities and satisfies all four axiomatic requirements we deem mandatory for this cost-sharing setting.

We then formulate the joint problem of partitioning users into subscription sharing groups and assigning cost-optimal data plans to them. Since the achievable savings in subscription costs depend closely on how well the derived subscription groups match with the shared data plan offer, it is plausible to address the two tasks simultaneously rather than in isolation. We characterize the complexity of the joint problem and devise three heuristic polynomial-time algorithms that leverage clustering techniques to solve it. The first one, called Agglomerative Cost-Minimization Clustering (ACMC), addresses the two parts of the joint problem simultaneously. At each step, it determines both the clusters that will be merged *and* the data plan that will be (tentatively) assigned to the new cluster to maximize the aggregate subscription cost savings. The other two algorithms decompose the problem into its two parts and solve them sequentially. Both of them try to cluster users so that the fluctuation of the collective data demand across the charging periods is minimized. The underlying idea is that subscription savings can be achieved by grouping together users with data demands that "cancel out" across different charging periods (*i.e.*, when one user consumes above the average in charging period x, the other(s) consume below the average and vice versa during another charging period y).

Finally, we show that important savings in subscription charges are achievable when sharing is facilitated for open groups of users, even under the worst of the three algorithms. These savings range from 20% up to 80% of what users would pay with cost-optimal individual data plans. Most notably, they turn out to be robust to the inaccurate prediction of the user's data consumption and the elasticity that commonly characterizes the user's demand for (cellular) data.

We present our models for the users' data demand and the shared capped data plans, including the DPCS rule, in section II. The joint optimization problem is formulated and characterized in section III and the three clustering-based algorithms are described in section IV. We evaluate their performance in section V under both perfect and imperfect prediction of users' data consumption and in section VI when this consumption adapts to the data plan cap. We conclude the paper with some thoughts as to how cellular mobile network operators could approach shared data plans.

II. MODELING USERS' DATA CONSUMPTION AND (SHARED) CAPPED DATA PLANS

The focus of our work is on capped data plans that stand in offer in the mobile data market. Capturing how users' data consumption is affected by the caps each data plan introduces is not a trivial problem. Both intuition and experience suggest that data consumption is elastic, *i.e.*, users adapt their data consumption patterns to the provisions of the plan (cap,

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2022.3231840

overage charges) they subscribe to. As a result, the actual consumption (realized demand) is part of the *a priori* intended demand for data, much as in the airlines' industry the realized bookings in a fully-booked flight forms the censored demand, a part of the *actual* demand for the flight (see *e.g.*, [13]).

Nevertheless, there is no consensus in the existing literature as to how this demand elasticity should be captured. At one extreme, in [8], the authors come up with a detailed model of how a fully rational and strategically acting user optimizes her daily consumption depending on the residual data cap and the instantaneous utility that data consumption bears. At the other extreme, in [4], the user suppresses a fixed portion of her *a priori* demand for data over a given charging period, if this exceeds the data plan cap.

In this work, we make the assumption that each user $u \in \mathcal{U}$, with $|\mathcal{U}| = U$, is described by a demand profile $\{d_{um}\}, m \in \{1, 2, ..., T\}$, where T is the number of charging periods (typically, the charging period is taken to be a month and T equals 12). This profile is essentially an *estimate* of the user's expected monthly data consumption and may take into account different types of information [11]. Much of this information can be provided by the user upon her registration with the platform and includes records of her past data monthly consumption as well as responses to an entry questionnaire about her interests and Internet usage patterns. Other information of interest are the trends of mobile data consumption growth at national or more local level and within different social and professional groups. In any case, these profiles, which can be updated over time, are communicated to the user and set an important reference for her data consumption. They are used for distributing the monthly charge of the shared data plan between its subscribers (see section II-B) and clustering users into subscription groups (see section IV).

A. Capped data plans

A capped data plan $p = (c_p, f_p, e_p)$ typically comes with a consumption cap c_p, monthly fee f_p and a penalty fee rate e_p in €/MB for excess consumption beyond the monthly cap. As a second option, instead of charging a fixed penalty rate per MB of excess consumption, some operators offer supplement data packages plans of c_e MB at fixed price f_e. These packages are not sold separately but only in conjunction with a "main" data plan. In either case, unused capacity during one charging period is not transferable to the next one. We denote with $\mathcal{P}$ the set of all individual plans that are available as subscription options to users, with $P = |\mathcal{P}|$. The notation used in the remainder of the paper is summarized in Table I.

Formally data plans are cost functions $C(q)$ of consumed data q. The two types of functions corresponding to the two main data plan options are:

$$C_1(q) = f_p + \max(0, q - c_p) \cdot e_p \quad \text{and} \tag{1}$$

$$C_2(q) = f_p + f_e \cdot \lceil \max(0, (q - c_p)/c_e) \rceil \tag{2}$$

where $\lceil y \rceil$ rounds y to the smallest integer $z \geq y$. These two functions are shown in Fig. 1. Both are non-decreasing; $C_1(q)$, hereafter called type-1 data plan, is continuous, whereas

TABLE I
LIST OF NOTATIONS USED IN THE PAPER

Symbol	Meaning
$\mathcal{U}(U)$	set (number) of mobile cellular users
u	a mobile cellular user in $\mathcal{U}$
$\mathcal{P}(P)$	set (number) of data subscription plans
p	a cellular data plan in $\mathcal{P}$
$\mathcal{G}$	set of possible partitions of the set $\mathcal{U}$
G	one partition of $\mathcal{U}$ in $\mathcal{G}$
g	a subscription-sharing group of users
g_{max}	maximum allowable size of subscription-sharing group
T	number of charging periods
d_{um}	estimated demand of user u in charging period m (MB)
q_{um}	actual data consumption of u in charging period m (MB)
q_{gm}	actual data consumption of subscription group g in charging period m (MB)
y^{ξ}_{um}	share of user u in the overall fee charged by a shared data plan in charging period m under cost-sharing rule ξ
$C_p(q)$	cost of data plan p as a function of consumption q
c_p	data consumption cap for plan p
e_p	penalty rate charged by type-1 data plan p
f_p	fixed monthly fee charged by data plan p
f_e	price of supplement data package for type-2 plans
c_e	data capacity of supplement data package for type-2 plans
$s_u(p, g)$	monetary savings of user u under shared data subscription plan p while member of subscription group g compared to the best individual plan
$s_{u,n}(p, g)$	normalized monetary savings of user u under data subscription plan p while member of subscription group g
$S_{opt}(\pi)$	optimal solution of the (OPT) problem instance p
$S_{alg}(\pi)$	algorithm's alg solution of the (OPT) problem instance p
$r_{alg}(\pi)$	empirical approximation ratio alg for the (OPT) problem instance p

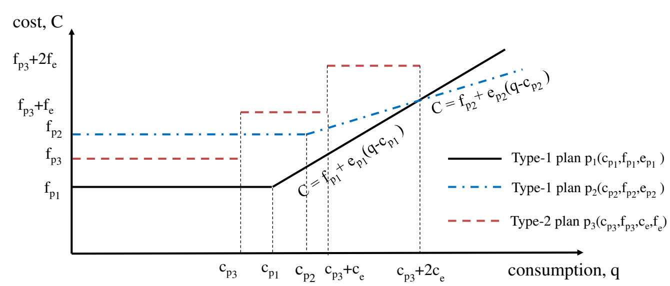


Fig. 1. Example data plans of type-1 (p_1 and p_2) and type-2 (p_3): $f_{p1} < f_{p2} < f_{p3}$, and $e_{p1} > e_{p2}$.

$C_2(q)$, hereafter called type-2 data plan, is piecewise constant. Neither of the two data plan types is differentiable.

Hence, if q_{um} is the amount of data that user u consumes during the m^{th} charging period, she is charged with $C_p(q_{um}) = C_1(q_{um})$ in case of a type-1 data plan, and $C_p(q_{um}) = C_2(q_{um})$ under a type-2 data plan.

In either case, she pays a total amount of

$$C_p(q_u) = \sum_{m=1}^{T} C_p(q_{um}) \tag{3}$$

over T charging periods.

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2022.3231840

B. Shared capped data plans

A shareable data plan p is described by the same three parameters, c_p, f_p, e_p (c_e, f_e for type-2 data plans).[1] If $g \subset \mathcal{U}$ is the group of users who share the data plan, the full charge that has to be paid by its members at charging period m is

$$C_p(q_{gm}) = f_p + \max(0, q_{gm} - c_p) \cdot e_p \tag{4}$$

for type-1 data plans and

$$C_p(q_{gm}) = f_p + f_e \cdot \lceil \max(0, (q_{gm} - c_p)/c_e) \rceil \tag{5}$$

for type-2 data plans, where $q_{gm} = \sum_{u \in g} q_{um}$ is the data consumption of the subscription (sharing) group g in the m^{th} charging period.

Then each group member u pays a share y_{um} of the overall subscription group's charge depending on her own consumption, the data consumption of the other group users, and the specific *cost-sharing scheme* that is used to split the overall plan cost into the members of the subscription group g. Formally, the cost-sharing scheme is a function $\xi_p(\cdot)$ that maps a vector of users' data consumption values $\{q_{um}\}_{u \in g}$ to cost shares $\{y_{um}\}$ for each user u in the sharing group g.

C. Sharing the capped data plan cost

1) Four requirements for the cost-sharing scheme ξ: We list four axiomatic requirements that a data plan cost-sharing scheme ξ should satisfy:

- (R1) It should always compute cost shares that sum exactly to the data plan cost, *i.e.*, $\sum_{u \in g} y_{um} = C_p(q_{gm})$. This cost includes both the monthly subscription fee f_p of the data plan and the penalty fee due to consumption beyond the data plan cap.
- (R2) It has to be symmetric in all user-related data consumption variables. Let $\{y_u\}_{u \in g}$ be the cost shares that ξ computes for a set $\{q_{um}, d_{um}\}_{u \in g}$ of values per group user. Then, for each arbitrary permutation of these values across the members of the sharing group, the cost shares that ξ computes should be the respective permutation of their cost shares $\{y_u\}_{u \in g}$. This ensures that ξ does not discriminate against any group user.
- (R3) The cost share that ξ computes for a given user should be a non-decreasing function of her own consumption. Irrespective of what the other members consume, it should not be possible for a user to increase her data consumption (thus, either leaving intact or increasing the overall data plan cost that is charged to the group) and, at the same time, reduce her own cost share.
- (R4) Finally, and less trivially, ξ should split the monthly fee f_p of the plan in proportion to users' contributions to the cost and it should not penalize a user with excess fees if she is not responsible for excess data consumption.

[1]The operators' current practice with data plans shared by family members or devices of the same user is to impose an additional fixed charge, o_p for each additional user/device sharing the plan. To the best of our understanding, this serves as a "penalty" fee, which compensates for the operator's revenue loss due to sharing. We do not consider this fixed fee further in our analysis.

The first requirement is a prerequisite for the efficiency and practical implementation of the subscription scheme. (R2)-(R4), on the other hand, collectively reflect that these schemes should be "fair" against all members of the sharing group.

2) The need for a new cost-sharing scheme: Cost-sharing schemes have been proposed in the economics and computer science area; see, for example, [14] and [15]. In what follows, we list the four most popular schemes and their prescriptions for the user cost shares.

(a) Average Cost Pricing (ACP): With ACP, the cost-share of each user is proportional to her consumption.

$$y_{um}^{ACP} = \frac{q_{um}}{q_{gm}} C(q_{gm}) \tag{6}$$

(b) Incremental Cost Sharing (ICS): With ICS, each user is charged for the marginal cost she generates when added to the subscription group.

$$y_{um}^{ICS} = C(q_{gm}) - C(\sum_{v \in g \setminus u} q_{vm}) \tag{7}$$

(c) Serial Cost Sharing (SCS): SCS demands that

$$y_{um}^{SCS} = \frac{1}{|g| - j + 1} C(q^j) - \sum_{k=1}^{j-1} \frac{1}{(|g| - k + 1)(|g| - k)} C(q^k) \tag{8}$$

where j is the index of user u when the group's user consumption values $\{q_{um}\}_{u \in g}$ are arranged in increasing order, $q_{1m} \leq q_{2m} \leq ... \leq q_{|g|m}$ and $q^j = (|g| - j + 1) q_{jm} + \sum_{k=1}^{j-1} q_{km}$.

(d) Shapley value (SV): Let σ be an arbitrary permutation of users in a subscription group g and $\sigma(k), 1 \leq k \leq |g|$ return the user in position k of the permutation. We can recursively compute the ordered marginal costs for all users in this permutation as

$$\begin{aligned} y_{\sigma(1)m}^{\sigma} &= C(q_{\sigma(1)m}) \\ y_{\sigma(i)m}^{\sigma} &= C(q_{\sigma(1)m}, q_{\sigma(2)m}, ..., q_{\sigma(i)m}) - \sum_{l=1}^{i-1} y_{\sigma(l)m}^{\sigma} \end{aligned}$$

Then the Shapley value for user u will be equal to the expectation of her ordered marginal costs

$$y_{um}^{SV} = \mathbb{E}(y_{um}^{\sigma}) \tag{9}$$

over all possible permutations of the subscription group members, which are assumed to be equiprobable.

We can show that

Proposition 1. *None of the four cost-sharing schemes, ACP, ICS, SCS or SV, satisfies all four requirements (R1)-(R4) for the cost functions (4) and (5).*

Proof. All four schemes trivially satisfy (R2) and (R3) and three of them, namely ACP, SCS and SV, also satisfy (R1). ICS fails, at least, (R1): it computes zero cost shares for all users of a group as long as the sum of their monthly data consumption values does not exceed the data plan cap.

ACP and SCS fail (R4) in different ways. ACS does so when the overall group consumption exceeds the data plan cap. Then, ACP shares the penalty fee between all users, even when the cap is exceeded because only one user consumes

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2022.3231840

aggressively. To see this, consider two users A and B, whose monthly consumption amounts, q_A and q_B, originally sum up exactly to the data plan cap c_p. Then, one of the two users, say A, increases her consumption by Δq and generates a penalty fee $\Delta q \cdot e_p$. The new cost share of user B is

$$y'_B = \frac{q_B}{q_A + q_B + \Delta q}(f_p + \Delta q \cdot e_p) > \frac{q_B}{q_A + q_B} f_p \qquad (10)$$

and exceeds her fixed fee share as long as

$$e_p > \frac{f_p}{c_p} \qquad (11)$$

namely, when the penalty rate exceeds the average cost (per MB) of the data plan cap, which is (almost) always the case. Hence, user B undertakes part of the excess fee due to the aggressive consumption by user A.

On the other hand, it takes a few more algebraic computations to show that, when there is no excess consumption, SCS splits the monthly subscription fee f_p of the data plan equally between all the users of the subscription group, irrespective of their individual data consumption levels.

Finally, SV fails in both ways. When there is no excess consumption, it computes equal cost shares for all the group subscribers, independently of how they individually contribute to the overall consumption. To see this, there are $|g|!$ permutations for each shared subscription group and the ordered marginal costs of user u are non-zero and equal to f_p for the $(|g|-1)!$ permutations that feature her in the leading position. Hence, her SV share is $f_p \cdot (|g|-1)!/|g|! = f_p/|g|$. Likewise, under excess consumption by a single group member, we can easily construct examples where the excess cost is spread across all group users. □

3) Double proportional cost sharing: We draw on the ACP scheme to come up with what we call the *double proportional cost sharing (DPCS) scheme*, shown in Algorithm 1. As its name suggests, DPCS applies the proportional rule of ACP twice, once to the fixed monthly fee (line 2 in Algorithm 1) and, potentially, a second time, to the penalty fees that result from excess consumption at group level (line 5). In the first case, the weights used in the proportional sharing correspond to the profile demands of the subscription group's members, whereas, in the second case, as weight for each subscriber stands the difference between its actual consumption and its quota of the shared data cap (line 4). Only users who exceed their quota contribute to the penalty fee of the group; a subscriber that sticks to her quota does not pay any excess fee. On the other hand, even if a user u consumes less than its precise quota $c_p \frac{d_{um}}{\sum_{u \in g} d_{um}}$ during charging period m, she is still charged her share of the fixed subscription fee (line 2). This is standard practice with individual capped plans as well, where the user is charged a fixed fee even in the extreme case that she does not consume *any data*. In shared plans, this also serves as a preventive measure against the unfair penalization of active group users due to one or more non-active ones, without whom they could have subscribed to a data plan with smaller cap.

Algorithm 1 Implementation of the DPCS scheme for type-1 data plans.

Input: User data consumption vector $\{q_{um}\}$ and profile demand vector $\{d_{um}\}$, $u \in g$, data plan $p = (f_p, c_p, e_p, o_p)$
Output: Individual cost shares $\{y_{um}\}, u \in g$

1: **for** user u in g **do**
2: $\quad y_{um} \longleftarrow f_p \frac{d_{um}}{\sum_{u \in g} d_{um}}$
3: $\quad$ **if** $\sum_{u \in g} q_{um} > c_p$ **then**
4: $\quad\quad excData(u) \longleftarrow \max(0, q_{um} - c_p \frac{d_{um}}{\sum_{u \in g} d_{um}})$
5: $\quad\quad y_{um} \longleftarrow y_{um} + \frac{excData(u)}{\sum_{u \in g} excData(u)} \cdot (C_p(\sum_{u \in g} q_{um}) - f_p)$
6: $\quad$ **end if**
7: **end for**
8: **return** $\{y_{um}\}$

Proposition 2. *The DPCS cost-sharing rule satisfies all four requirements (R1)-(R4) for the cost functions (4) and (5).*

Proof. The scheme trivially satisfies (R1) since both the fixed and the excess costs are partitioned into subscriber shares that add up to the full costs (ref. lines 2 and 5 in Algorithm 1). It satisfies (R2) since the variables $\{q_{um}, d_{um}\}$ suffice to fully determine the cost shares of the subscription group members and no user discrimination is possible by other means. (R3) also holds in that as the consumption of a subscriber increases, the cost it has to pay either remains the same (if her monthly quota is not exceeded) or increases (if it exceeds its quota and the excess cost is non-zero). Finally, (R4) is met by design in the DPCS rule. The scheme was devised to exactly respond to this requirement, which is not met by the other three cost-sharing rules in Proposition 1.

□

From Algorithm 1, it is clear that the savings end users can achieve through data plan sharing depend on how they are organized into subscription groups. This point is further reinforced by the toy example in an earlier version of our work [16]. In what follows, we formulate this dependency into an optimization problem.

III. The Joint User Partitioning and Data Plan Assignment Problem

A. Problem formulation

For a given offer of cellular data plans $\mathcal{P}$, with $P = |\mathcal{P}|$, the mission of the platform is to partition users into subscription groups and assign those groups to data plans such that their achieved savings in data plan subscription fees are maximized. The two tasks are interrelated: the savings that can be achieved for each subscription group and its members with the best available plans depend on the actual subscription groups that will be formed (*i.e.*, the way individual data consumption patterns are aggregated).

Let G be a user partition and $g_1, g_2, ..., g_n$ the subscription groups that make it up with $|g_i| \leq g_{max}, 1 \leq i \leq n$, $g_i \bigcap g_j = \emptyset, \quad \forall g_i, g_j \in G$ and $\bigcup_{i=1}^{n} g_i = \mathcal{U}$. The monetary savings

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2022.3231840

of a user u, when she subscribes to plan p as member of subscription group g are:

$$s_u(p,g) = \sum_{m=1}^{T} \left(C_{p_i}(q_{um}) - \xi_{pg}(q_{um}, \mathbf{q}_{-um})\right) \quad (12)$$

where $\mathbf{q}_{-um} \coloneqq \{q_{um}\}_{u \in g \setminus u}$, is the vector of the consumed data amounts by the other group members, $\xi_{pg}(q_{um}, \mathbf{q}_{-um})$ her cost share under plan p during charging period m and

$$p_i = \arg\min_{p \in \mathcal{P}} \sum_{m=1}^{T} C_p(q_{um}) \quad (13)$$

is the individual (non-shared) plan that minimizes what the user pays for given consumption if she subscribes to it alone.

Since any given subscription savings are of higher value to someone paying low subscription fees rather than to someone paying high fees, we normalize those savings with respect to what users pay under the most cost-effective individual plan

$$\begin{aligned} s_{u,n}(p,g) &= \frac{s_u(p,g)}{\sum_{m=1}^{T} C_{p_i}(q_{um})} = 1 - \frac{\sum_{m=1}^{T} \xi_{pg}(q_{um}, \mathbf{q}_{-um})}{\sum_{m=1}^{T} C_{p_i}(q_{um})} \\ &= 1 - r_u \cdot \sum_{m=1}^{T} \xi_{pg}(q_{um}, \mathbf{q}_{-um}) \\ &= 1 - c_{u,n}(p,g) \end{aligned} \quad (14)$$

where $r_u = 1/\sum_{m=1}^{T} C_{p_i}(q_{um})$.

The platform then seeks to partition users into subscription groups and assign to them data plans that maximize their normalized cost savings; or, equivalently, according to (14), data plans that minimize their cost shares, weighted by the cost of the best individual plan for the same consumption. Formally, the platform is after an optimal partition G^* of users among the set of all feasible partitions $\mathcal{G}$ and an assignment $\mathbf{x}^* = \{x_{pg}\}$ of data plans $p \in \mathcal{P}$ to the subscription groups $g \in G^*$, with $x_{pg} = 1$ if data plan p is assigned to group g and $x_{pg} = 0$, otherwise, that solve the optimization problem

$$\min_{\mathbf{x},\mathcal{G}} \quad \sum_{g \in G} \sum_{u \in g} \sum_{p \in \mathcal{P}} c_{u,n}(p,g) x_{pg} \qquad (OPT) \quad (15)$$

$$s.t. \quad 1 \le |g| \le g_{max}, \quad g \in G \quad (16)$$

$$\sum_{p \in \mathcal{P}} x_{pg} = 1, \quad g \in G \quad (17)$$

$$0 \le c_{u,n} \le 1, \quad u \in \mathcal{U} \quad (18)$$

$$x_{pg} \in \{0,1\}, \quad g \in G, G \in \mathcal{G}, p \in \mathcal{P} \quad (19)$$

In OPT, constraint (16) enforces that the group sizes in each acceptable partition $G \in \mathcal{G}$ are upped bounded by g_{max}. The equality (17) reflects that one plan is assigned to each such group, (18) ensures that no user experiences loss through plan sharing compared to what she would pay under the best individual plan and (19) explicates that the assignment variables are binary. Note that OPT does not preclude singleton subscription groups, *i.e.*, instances that no {subscription group, shared data plan} pair is more economical for a given user than a "normal" non-shared plan. Moreover, it readily generalizes to cost-sharing rules other than DPCS. It suffices to compute the respective terms $\{c_{u,n}\}$ in (15) according to (14).

B. Problem characterization

The problem OPT is a joint (user) partitioning and (data plan) assignment problem. The number of possible partitions of users into subscription groups of size up to g_{max} is given by the restricted Bell numbers $B_{U \le g_{max}}$ [17] [18]. These numbers are computed through the recursion

$$B_{n \le m} = \sum_{k=0}^{m-1} \binom{n}{k} B_{n-k-1 \le m} \quad (20)$$

with $B_{n \le m} = B_n$ for $n \le m$, where B_n is the Bell number denoting the number of partitions of an n-element set into parts of arbitrary size. The Bell numbers, are also recursively computed:

$$B_n = \sum_{k=0}^{n-1} \binom{n-1}{k} B_k \quad (21)$$

with $B_0 = 1$, $B_1 = 1$ and $B_2 = 2$. The Bell numbers grow exponentially fast with n and so do the restricted Bell numbers.

We can show that

Proposition 3. *The problem OPT is NP-hard.*

Proof. Consider the special case of the problem, say OPT_s, where $P = 1$, *i.e.*, a single mobile data plan is at offer. Then the joint partitioning and data plan assignment problem reduces to a partitioning problem. OPT_s can be identified as an instance of the set-partition problem (SPP) with an additive objective, see [19]. More specifically, the SPP with an additive objective reduces to OPT_s if (a) users and subscription groups in OPT_s are mapped to the items and subsets of items, respectively, in the general statement of the SPP (see, for instance, [20]); (b) the per-user costs to be minimized in (15) are summed over the users of each subscription group $g_i \in \mathcal{G}$ to yield subset-specific costs in the SPP formulation.

The SPP with an additive objective is shown to be NP-hard in [19] for arbitrary cost functions under the condition that the problem accepts partitions with U subsets. In OPT_s this condition is trivially satisfied since one feasible solution with U subscription groups corresponds to all users sticking to their most cost-effective *individual* plans without extracting any benefits from plan sharing. Hence, OPT_s is also NP-hard and, by a generalization argument, OPT is NP-hard as well. □

IV. Solving the Joint User Partitioning and Data Plan Assignment Problem

We present three algorithms for the joint problem. They leverage clustering starting with singleton clusters for individual users and work with their demand profiles, $\{d_{um}\}_{u \in U}$. The first one attacks the user partitioning and data plan assignment tasks simultaneously, whereas the other two decompose the problem: first, they partition users into subscription groups and then, in a second simpler step, they identify optimal shared data plans for them.

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2022.3231840

Algorithm 2 Agglomerative cost-minimization clustering

Input: User demand profiles $\{d_u\}, u \in \mathcal{U}$; group size limit, g_{max}; data plan cost functions $C_p(q), p \in \mathcal{P}$
Output: Sharing groups, $\{g\}$, and data plan assignments, $p_{opt}(g)$, $\cup g = \mathcal{U}$

Initialization step
Start with one cluster for each user: $g_u \longleftarrow u \in \mathcal{U}$
Compute $p_{opt}(u) \longleftarrow \min_{p\in\mathcal{P}} C_p(D_u), \forall u \in \mathcal{U}$
while there are clusters with size $< g_{max}$ and merging is possible **do**
  User partitioning step
  For each cluster pair (g_k, g_l) compute score(g_k, g_l) from (22), (23)
  Merge the two clusters $(g_{k'}, g_{l'})$ with the highest positive score
  Data plan assignment step
  $p_{opt}(g_{k'} \cup g_{l'}) \longleftarrow \arg\min_{p\in\mathcal{P}} \sum_{m=1}^{T} C_p(\sum_{u\in g_{k'}\cup g_{l'}} d_{um})$
end while

A. Agglomerative cost-minimization clustering (ACMC)

The algorithm first identifies the optimal data plan for each user, *i.e.*, the plan that minimizes her expected charge over T charging periods under her demand profile. It then initiates an agglomerative clustering process. At each step in this process, the algorithm merges those clusters g_k, g_l, with $|g_k| + |g_l| \leq g_{max}$ that maximize the normalized subscription cost savings for the members of the two clusters

$$score(g_k, g_l) = \frac{\sum_{m=1}^{T}\left(C_{p_k}(d_{g_k m}) + C_{p_l}(d_{g_l m}) - C_{p_{kl}}(d_{g_{kl} m})\right)}{\sum_{m=1}^{T}\left(C_{p_k}(\sum_{u\in g_k} d_{um}) + C_{p_l}(\sum_{u\in g_l} d_{um})\right)} \tag{22}$$

under the assumption that cost-optimal data plans

$$p_k = \arg\min_{p\in\mathcal{P}} \sum_{m=1}^{T} C_p(\sum_{u\in g_k} d_{um}) \quad \text{and} \tag{23}$$

$$p_{kl} = \arg\min_{p\in\mathcal{P}} \sum_{m=1}^{T} C_p(\sum_{u\in g_k \cup g_l} d_{um}) \tag{24}$$

are chosen in each case.

The clustering process ends when either the subscription group limit g_{max} is reached for each cluster or no further cost savings are possible for any subscription group. The pseudocode of the algorithm is shown in Algorithm 2.

Algorithm complexity: From a complexity point of view, the initialization step requires $O(UPT)$ steps, each involving a call in (1). The algorithm then carries out $O(U)$ steps and in each one of those, it searches for the best one out of $O(U^2)$ pairs of clusters to merge. For each one of the candidate cluster-pairs, the algorithm computes the minimum-cost plan, which requires $O(PT)$ calls to (1), plus a find-minimum operation over the candidate pairs that requires $O(U^2)$ time. Hence, if β is the time needed for one iteration of (1), the overall time-complexity of the algorithm is $O(U^3\beta PT)$.

B. Agglomerative uniform-consumption clustering (AUCC)

The ACMC algorithm simultaneously constructs subscription groups and assigns optimal data plans to them. On the contrary, the agglomerative uniform-consumption clustering algorithm decomposes the problem into its two subproblems: first, it clusters users into subscription groups, and, in a second step, it identifies optimal plans for them.

The metric that scores clusters throughout the process is the *normalized fluctuation of the group demand* over the period for which the users' demand profiles are available. Therefore, the demand fluctuation for a cluster g is measured by

$$d_F(g) = \frac{\max_{m=1}^{T} \sum_{u\in g} d_{um} - \min_{m=1}^{T} \sum_{u\in g} d_{um}}{\min_{m=1}^{T} \sum_{u\in g} d_{um}} \tag{25}$$

and in each step of the algorithm's execution, we merge existing clusters (g'_k, g'_l) such that

$$(g'_k, g'_l) = \arg\min_{g_k, g_l} d_F(g_k \cup g_l) \tag{26}$$

The intuition behind the cluster score is that cost savings are achieved when we group together users who can absorb the temporal fluctuation in each others' demands and together present a flatter profile that can be more easily matched to a data plan. This is reminiscent of the statistical multiplexing gains achieved in dimensioning telecommunication networks, when aggregating smaller traffic streams from end user access links to traffic aggregates in higher capacity links [21].

The algorithm terminates when either g_{max} is reached for each cluster or no further improvement is feasible in the normalized fluctuation of demand for any cluster pair during a merging step. Finally, the data plan p_k assigned to a group g_k is given by (23).

Algorithm complexity: Each merging step of the clustering algorithm takes $O(U^2)+O(U)) = O(U^2)$ time since we need to compute the score d_F for all possible cluster pairs and merge the two that minimize it. Overall, the time complexity of the algorithm is $O(U^3)$ for the clustering part, plus $O(UPT\beta)$ for the data plan assignment part, *i.e.*, $O(U^3)$ overall.

C. Double greedy maximal uniform-consumption clustering (DGMC)

Similar to the AUCC algorithm, this algorithm decomposes the original problem into the user grouping and data plan assignment subproblems and uses the normalized fluctuation of demand measure, d_F, to score clusters. However, the algorithm is no longer agglomerative. It rather searches iteratively and more exhaustively for possible subscription groups.

The algorithm starts from each individual user u and greedily builds U different u-maximal clusters, that is clusters that containing u and are maximal in the sense that they cannot increase any more either because their size is g_{max} or because no addition of another user can further decrease the

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2022.3231840

TABLE II
DATA PLANS USED FOR THE EVALUATION OF THE HEURISTIC ALGORITHMS: DATA CAPS (c_p), FIXED FEES (f_p) AND OVERAGE CHARGES (e_p).

ID	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}	p_{11}	p_{12}	p_{13}	p_{14}	p_{15}	p_{16}	p_{17}
c_p (GB)	0.5	2	5	10	20	100	0.4	1	2	4	5	10	1	3	7	15	30
f_p (€)	4.85	9.8	14.76	19.71	24.66	44.47	3.84	7.72	11.6	15.48	17.42	23.24	4.99	9.99	19.99	29.99	49.99
e_p (€/MB)	0.019	0.019	0.019	0.019	0.019	0.019	0.039	0.039	0.039	0.039	0.039	0.039	0.19	0.19	0.19	0.19	0.19

Algorithm 3 Agglomerative uniform-consumption clustering

Input: User demand profiles $\{d_u\}, u \in \mathcal{U}$; group size limit, g_{max}; data plan cost functions $C_p(q), p \in \mathcal{P}$
Output: Sharing groups, $\{g\}$, and data plan assignments, $p_{opt}(g)$, $\cup g = \mathcal{U}$

Initialization step
Start with one cluster for each user: $g_u \longleftarrow u \in \mathcal{U}$
User partitioning step
while there are clusters with size $< g_{max}$ and merging is possible **do**
  For each pair of clusters (g_k, g_l) compute score(g_k, g_l) after (25)
  Merge the two clusters $(g_{k'}, g_{l'})$ after (26)
end for
Data plan assignment step
for every group g in the resulting cluster structure **do**
  $p_{opt}(g) \longleftarrow \min_{p\in\mathcal{P}} \sum_{m=1}^{T} C_p(\sum_{u\in g} d_{um})$
end for

Algorithm 4 Double greedy maximal uniform-consumption clustering

Input: User demand profiles $\{d_u\}, u \in \mathcal{U}$; group size limit, g_{max}; data plan cost functions $C_p(q), p \in \mathcal{P}$'
Output: Sharing groups, $\{g\}$, and data plan assignments, $p_{opt}(g)$, $\cup g = \mathcal{U}$

Initialization step
Start with one cluster for each user, $g_u \longleftarrow u \in \mathcal{U}$
User partitioning step
while $\mathcal{U} \neq \emptyset$ **do**
  Build d_F-maximal clusters $cl(u), u \in \mathcal{U}$
  Rank the U maximal clusters $cl(u)$ in order of non-decreasing d_F, see (25).
  while $|\{cl(u)\}| > 1$ **do**
    Add the top disjoint clusters $\{m\}$ to the clustering structure
    $\mathcal{U} \longleftarrow \mathcal{U} \setminus \{u \in \{m\}\}$
  end while
end while
Data plan assignment step
for every group g in the resulting cluster structure **do**
  $p_{opt}(g) \longleftarrow \min_{p\in\mathcal{P}} C_p(\sum_{u\in g} d_{um})$
end for

cluster's d_F value. Since these U clusters typically overlap, the algorithm ranks the clusters in order of non-decreasing d_F and, greedily, retains as many as possible disjoint clusters.

Users who are included in those clusters are removed from consideration in the second iteration of the algorithm, which builds maximal clusters from scratch for the remaining users. A new set of disjoint clusters with minimum d_F scores is chosen and the corresponding users are removed from consideration. This doubly greedy process of maximal cluster formation and selection of disjoint clusters continues until all users are clustered. Note that some users may end up standalone if no pairing with another user can decrease the fluctuation in their aggregate consumption. The resulting groups are then matched, in a subsequent step, with the shared data plans that minimize the subscription fees they need to pay as groups. The overall algorithm is shown in Algorithm 4.

Algorithm complexity The first iteration of the algorithm includes all $U_1 \equiv U$ users and requires $O(U_1^{g_{max}})$ steps for building maximal clusters plus $O(U_1 ln U_1)$ time for sorting the clusters and picking the maximum possible number of disjoint ones. Subsequent iterations involve reduced sets of users $U_k < U$ and require time $O(U_k^{g_{max}} + O(U_k ln U_k)$. The overall time for the clustering step is $\sum_k \left(O(U_k^{g_{max}}) + O(U_k ln U_k) \right) \subseteq O(U^{1+g_{max}})$. An additional $O(UP\beta)$ time is needed for the data plan assignment step.

V. EVALUATING THE THREE ALGORITHMS

We first compare the solutions of the three clustering-based algorithms with the optimal one. This comparison is only feasible for small problem instances that do not render prohibitive the computation of the optimal solution through exhaustive enumeration[2]. Then, we assess the savings the three algorithms achieve, the sizes of subscription groups they generate and how tight are the caps of the assigned plans with respect to the groups' demand profiles. Finally, we explore the sensitivity of the three algorithms to the prediction accuracy of the user demand profiles.

A. Comparison with optimal solution

If $S_{alg}(\pi)$ are the normalized savings achieved by algorithm $alg \in \{'ACMC', 'AUCC', 'DGMC'\}$ for an instance π of the (OPT) problem, and $S_{opt}(\pi)$ the optimal solution, the empirical approximation ratio $r_{alg}(\pi)$ is defined as the ratio

$$r_{alg}(\pi) = \frac{S_{alg}(\pi)}{S_{opt}(\pi)} \tag{27}$$

[2]We enumerate all possible partitions of users into subscription groups and for each group we identify the data plan that results in highest normalized savings. The optimal partition is the one that minimizes the aggregate savings over all users.

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2022.3231840

TABLE III
EMPIRICAL APPROXIMATION RATIOS OF THE THREE ALGORITHMS FOR SMALL (U, g_{max}) VALUES

U	g_{max}	runs	Time	r_{DGMC}	r_{AUCC}	r_{ACMC}
9	2	400	1H17m	0.916	0.91	0.98
9	3	100	1H	0.849	0.92	0.96
9	4	70	1H6m	0.817	0.81	0.94
9	5	60	1H10m	0.761	0.89	0.97
10	2	150	1H07m	0.946	0.94	0.95
10	3	30	1H16m	0.89	0.88	0.95
10	4	14	1H45m	0.762	0.81	0.94
10	5	12	1H15m	0.772	0.89	0.97
11	2	60	1H51m	0.918	0.92	0.99
11	3	3	2H12m	0.953	0.97	0.96
11	4	1	1H28m	0.76	0.86	1.0

Table III reports the average empirical approximation ratio, computed for given (U, g_{max}) pairs of values, over a number of runs. The runs for each pair of values terminate when one hour is exceeded, *i.e.*, the last run is the one that pushes the overall experimentation time for the given (U,g_{max}) beyond the 1hr threshold. In each run, we randomly generate U user demand profiles and consider the cellular data plans in Table II. We then generate user partitions through enumeration, each time using one of the three algorithms, and assign each of them to the data plan that maximizes savings for its subscribers.

Although the sample of problem instances is small, we can note that (a) ACMC performs distinctly better with empirical approximation ratios that exceed 0.94; (b) the AUCC algorithm is the second best with empirical approximation ratio scores in the range [0.81,0.92] and (c) DGMC performs worse, featuring the lowest average values and the highest variance.

B. Performance of the algorithms

1) Methodology: The main building blocks for the experimental comparison of the three algorithms are:

User demand profiles: Each user is represented by a synthetically generated T-dimensional profile demand vector d_u = [d_{u1}, d_{u2}, ..., d_{uT}], where T is the number of charging periods covered by the profile of u. We fix the average user monthly demand values $\overline{d_u}$ according to the mobile data plan distributions reported in [22] and generate the T demand values by sampling normal distributions $\mathcal{N}(\overline{d_u}, \sigma_u)$. Unless otherwise stated, the demand values are in MB, $T = 12$ and $\sigma_u = 0.2 \cdot \overline{d_u},\ u \in \mathcal{U}$ with $U = 1400$.

Cellular mobile data plans: We have identified and collected information about 17 different cellular data plans on offer by operators in various European countries. These plans are listed in Table II and make up the set $\mathcal{P}$ in the problem definition (see section II-A). They feature various data caps (400MB-100GB), monthly fees (4.85-49.99€) and overage charges.

Performance metrics: We compare the three algorithms along various dimensions. The ultimate performance measure are the cost savings they achieve for the mobile users. They are measured for each user both in absolute terms (in €)

$$sav(u) = \min_{p\in\mathcal{P}} \sum_{m=1}^{T} C_p(d_{um}) - \min_{p\in\mathcal{P}} \sum_{m=1}^{T} \xi_p(d_{um}, d_{-um}) \quad (28)$$

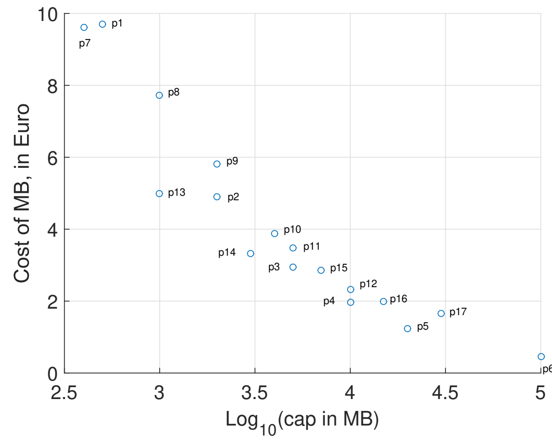


Fig. 2. Data plan cap *vs.* cost per MB for each of the 17 cellular data plans.

and in relative terms, *i.e.*, as their ratios over the charges under the optimal individual plan

$$nsav(u) = \frac{sav(u)}{\min_{p\in\mathcal{P}} C_p(\sum_{m=1}^{T} d_{um})} \quad (29)$$

We report histograms and empirical cumulative distribution functions of these savings over the user population. We also compute the portions of users who experience normalized savings beyond $\alpha \ \in [0, 1]$ as

$$perc(\alpha) = \frac{\sum_{u\in\mathcal{U}} 1_{nsav(u)>\alpha}}{U} \quad (30)$$

where 1_x is the indicator function that equals one when condition x is true.

Moreover, a number of statistics yield further insights into the way the three algorithms assign users to subscription groups. The first one is the *distribution of subscription group sizes* each algorithm generates. A second one relates to how well each subscription group utilizes the data plan it is assigned to, *i.e.*, how much data remain unused and how much excess consumption takes place.

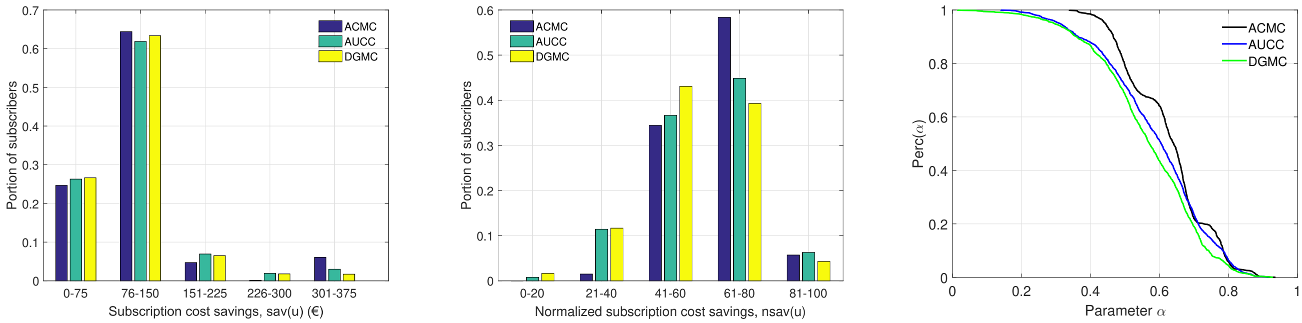
2) Results: The main results out of this comparison are summarized below:

Subscription cost savings: Fig. 3 reports the predicted cost savings per user, according to (28), (29) and (30), when the three algorithms derive subscription groups and assign data plans to them according to the user demand profiles.

The savings with the three algorithms appear to be comparable in absolute terms. Yet, the ACMC algorithm distinguishes from the other two in securing higher annual subscription savings, beyond 300€, for distinctly more subscribers (more than 6%) than the other two algorithms. This results in aggregate savings that are 13% (10%) higher than the DGMC (AUCC) algorithms, as shown in the last column of Table IV.

The performance advantage of the ACMC algorithm is more evident in terms of normalized savings, the measure that this algorithm actually tries to optimize (see Algorithm 2). The algorithm consistently tends to produce subscription groups that save more with respect to what their members pay under

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2022.3231840



(a) Absolute savings (b) Relative savings (c) Empirical ccdf of relative savings

Fig. 3. Distribution of per user subscription cost savings under the three algorithms, computed based on their demand profiles : $U = 1400$.

TABLE IV
DISTRIBUTION OF (NORMALIZED) SUBSCRIPTION COST SAVINGS UNDER PERFECT ESTIMATION OF USERS' DATA CONSUMPTION ($U = 1400$)

	absolute savings, sav(u) in €				relative savings, nsav(u)				$\sum sav(u)$
	[0,87]	[88,175]	[176,263]	> 263	[0,0.25]	[0.26,0.5]	[0.51,0.75]	[0.76,1]	
DGMC	40.78%	51.57%	5.21%	2.21%	3%	28.5%	59.71%	8.57%	137061.27€
AUCC	41.07%	49.35%	6%	3.57%	2.07%	26.21%	57.35%	14.35%	142971.3€
ACMC	26.85%	63.5%	3.42%	6.21%	0%	20.92%	60.35%	18.71%	157502.1€

TABLE V
GROUP SIZE DISTRIBUTION

	1	2	3	4	5
DGMC	0%	2.94%	2.52%	2.94%	91.5%
AUCC	0%	0.43%	0.43%	0%	99.13%
ACMC	0%	0.36%	17.71%	39.11%	42.8%

individual plans. This trend is clearer in the third plot of Fig. 3, where $\forall \alpha \in [0, 1]$

$$perc^{ACMC}(\alpha) > max\big(perc^{AUCC}(\alpha), perc^{DGMC}(\alpha)\big)$$

implying a stochastic dominance relationship of ACMC over the other two algorithms in terms of achievable normalized subscription savings.

Subscription group size: Table V yields more insights to the way the three algorithms work. Using the monthly fluctuation of user consumption as a proxy measure for their user-grouping decisions, AUCC and DGMC are strongly biased towards large subscription groups: AUCC gathers almost all users (99.13%) into maximum size subscription groups, while the respective number is around 10% smaller for the DGMC. On the contrary, the ACMC spreads users into subscription groups of size three to five in more balanced manner. Although larger subscription groups reduce the subscription charges leveraging the economy of scale properties of data plans (see Fig. 2), they do not necessarily do this in the optimal manner. Simultaneously solving the subscriber grouping and the data plan assignment tasks, the ACMC algorithm reaches better decisions about the number and sizes of subscription groups that maximize savings for the subscribers.

Fig. 4 suggests that the distributions of normalized subscription cost savings as well as the way the three algorithms compare with each other in this respect are practically independent of the number of subscribers. Fundamental descriptors of the distributions such as the location of the mode of the distribution and the ranking of the five intervals in the x-axis regarding the probability mass they accumulate persist across all values of U. To statistically verify the graphical evidence, we apply the two-sample Kolmogorov-Smirnov test for distributional similarity [23] in all six possible pairs of distributions ($U = 200, 300, 400, 1400$) for each algorithm. The test does not reject the null hypothesis that all pairs of samples come from the same distribution, even at 1% significance level, with the single exception of the cost savings under the AUCC algorithm and $U = 300$.

Utilization of data plans: One main reason for the remarkable savings that are achievable with shared data plans is the high under-utilization of *individual* capped data plans. Namely, their coarse granularity results in many users subscribing to plans with caps that significantly exceed their needs and are far from exhausted in most charging periods. As shown in Fig. 5, this capacity waste exceeds 30% of the data plan cap for almost 50% of the individual data plans and 15% of the cap for almost all individual plans. Ideally, such mismatches should be as small as possible. To the extent that capped data plans imply commitments on behalf of the cellular network operator to satisfy the user's demand if this is actually realized, underutilized/overdimensioned data plans only increase uncertainty and complicate the resource planning process.

Fig. 5 reveals that the ACMC algorithm generates (subscription group, data plan) pairs, where the group consumption (as predicted by profile demands) matches far more tightly the data plan cap than the pairs generated by the other two algorithms. AUCC and DGMC give rise to significant amounts of unused data capacity. In fact, their shared data plans are more underutilized than the original individual data plans.

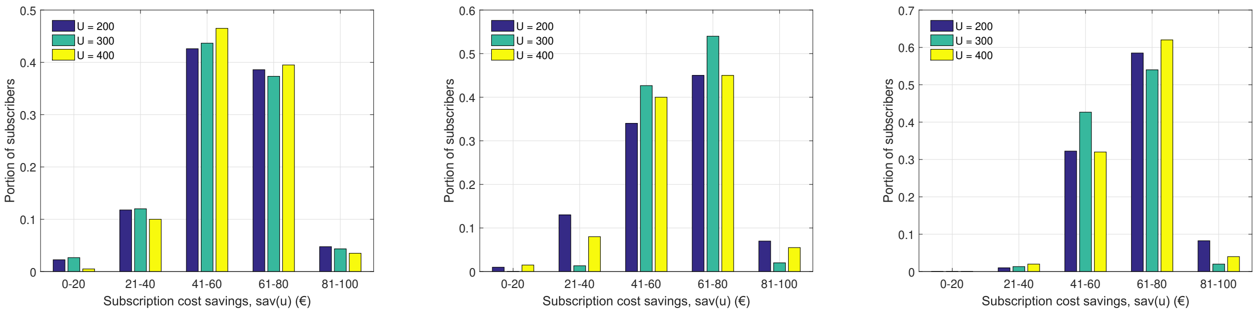
C. Sensitivity of savings to the prediction accuracy of users' data consumption

The grouping of cellular users into subscription groups and the assignment of shared data plans to them is carried out based on their demand profiles (see Algorithms 2-4), which can be extracted as discussed in section II. These profiles

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSNM.2022.3231840

TABLE VI
DISTRIBUTION OF (NORMALIZED) SUBSCRIPTION COST SAVINGS UNDER IMPERFECT PREDICTION OF USERS' DATA CONSUMPTION, $(\mu, \sigma) = (1.1, 0.12)$.

	absolute savings, sav(u) in €					relative savings, nsav(u)					$\sum sav(u)$
	< 0	[0,87]	[88,175]	[176,263]	> 263	< 0	[0,0.25]	[0.26,0.5]	[0.51,0.75]	[0.76,1]	
ACMC	1.21%	38.93%	52.43%	5.64%	1.79%	1.21%	9.07%	24.36%	52.21%	13.14%	136690€



(a) DGMC algorithm (b) AUCC algorithm (c) ACMC algorithm

Fig. 4. Distribution of per user normalized subscription cost savings, computed based on their demand profiles, under the three algorithms and variable number of cellular mobile users.

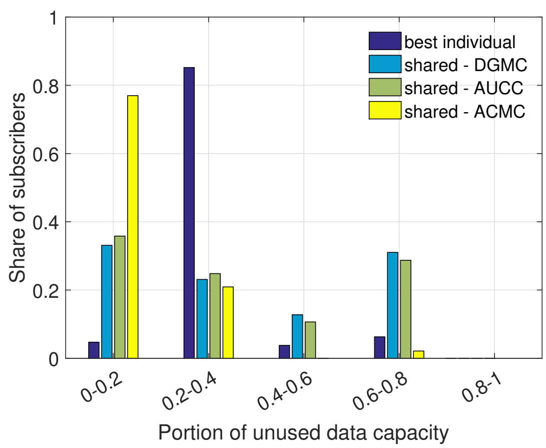


Fig. 5. Distribution of unused (wasted) data capacity under the best individual data plans and the shared data plans derived with the three algorithms

are *estimates* of the users' actual data consumption and the assumption behind the subscription savings reported in section V-B2, in Fig. 3 and Table IV, has been that these estimates are perfect. We now relax this assumption and question how the accuracy of those estimates affects the achievable subscription savings. We focus on ACMC, the best of the three algorithms according to our evaluation in section V-B2.

We recompute the savings through shared plans when the amount of data user u *actually* consumes each charging period $m \in \mathcal{T}$ is sampled from a normal distribution $\mathcal{N}(\mu \cdot d_{um}, \sigma \cdot d_{um})$. Hence, we let the mean actual user data consumption exceed the predicted one (profile) by a fixed factor mu (bias) and, on top of that, we let the consumption fluctuate around this mean. Parameter μ varies from 1.05 to 1.2 and σ varies from 0 (no variance) to 0.2, both in steps of 0.05.

Table VI reports the average savings that emerge when we simulate 100 realizations of actual users' data consumption, for $(\mu, \sigma) = (1.1, 0.12)$. Two are the immediate remarks when compared to Table IV. First of all, as expected, the aggregate savings achieved by the ACMC algorithm (last column) are 30-40% lower. Second, a small part of the subscribers (1.2%) now end up paying more with shared plans than they do under their best available individual plan (see the two columns in Table VI reporting negative savings). Overall, there is a visible shift of the savings' distribution towards smaller values.

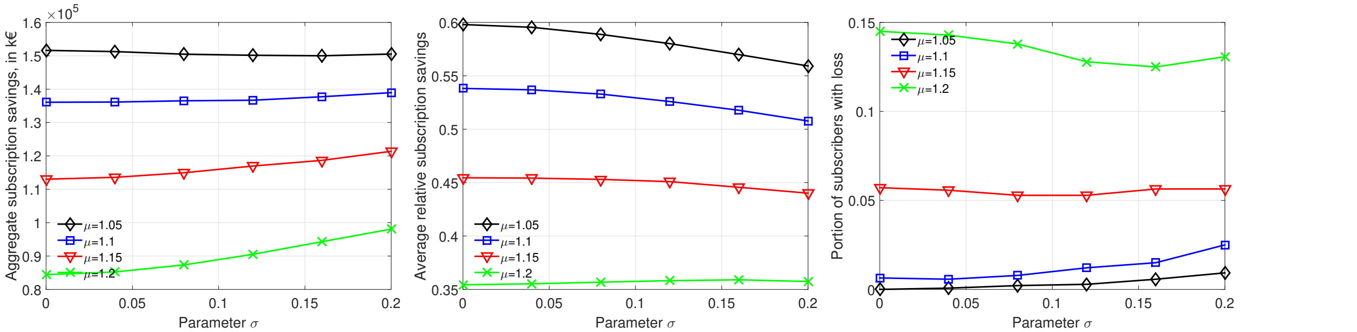
Then, Fig. 6 demonstrates how these findings generalize for other (μ, σ) value pairs. As the prediction bias grows, the subscription savings, both absolute and relative, almost halve and the number of subscribers that end up paying more than they do with the best individual plans grows from 1% up to 15%. On the contrary, the impact of fluctuations around the mean demand profiles is much less significant, as shown in Fig. 6a,b. Recall that ACMC finds data plans that suit more tightly the subscription groups it constructs. What marks a strong point under accurate demand profiles ends up being a weak point under inaccurate prediction since the data plan caps are exceeded more easily and penalty fees are charged more frequently. This further justifies the need for (periodic) re-estimation and adaptation of users' demand profiles, to ensure that the algorithm works with reasonably accurate estimates of users' actual data consumption.

VI. DEMAND ELASTICITY AND SHARED PLANS

So far we have assumed that the actual user consumption in each charging period is an independent parameter. The user demand for data is realized, irrespective of whether it can be accurately predicted or not. However, users tend to self-control their consumption, not least in response to warning messages from the mobile operator about approaching/exceeding their data quota. As a result, the actually realized consumption deviates from the originally *intended* demand.

We capture the elasticity of user demand through the model in [4]. Namely, we assume that when the consumption of a user group exceeds the data cap during a charging period, a fraction ϵ of the excess demand is not actually realized but it is rather suppressed by the user. The excess consumption cost is still distributed to the group members in line with the DPCS rule in Algorithm 1. The model directly aims at the net effect of the elasticity property and is much simpler than models that seek to describe in more detail the user behavior on a daily

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2022.3231840



(a) Aggregate absolute savings (b) Aggregate relative savings (c) Users with loss under sharing

Fig. 6. Subscription savings and users with loss under imperfect prediction of actual user data consumption: ACMC algorithm, $U = 1400$, each point in the plot is the average over 100 realizations of users' data consumption.

TABLE VII
NUMBER OF TOTAL SUBSCRIPTION GROUPS AND GROUPS THAT STAY THE SAME WHEN DEMAND ELASTICITY IS INTRODUCED.

ϵ	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
$\lvert G\rvert_\epsilon$	52	51	50	48	48	49	49	49	47
$\lvert G\rvert_\epsilon - \lvert G\rvert_0$	4	3	2	0	0	1	1	1	-1
unchanged	3	3	3	3	3	8	11	10	10

basis (*e.g.*, [8]). Nevertheless, it captures the dependence of the actual data consumption on the subscribed plan and helps us assess the impact of this demand elasticity on the achievable savings due to plan sharing.

We focus on the ACMC algorithm since ϵ does not only affect the choice of shared plan for each subscription group but it is also budgeted for by the algorithm while constructing subscription groups. We consider a sample of 200 users, who yield 48 subscription groups when $\epsilon = 0$ and data consumption is accurately predicted. We let ϵ vary in [0.1, 0.2, ..., 0.9]. Table VII reports the number of subscription groups $|G|_\epsilon$ that result for different ϵ values and compares them with their number $|G|_0$, when $\epsilon = 0$. The way users are organized into groups under the ACMC algorithm changes significantly. Although the overall number of subscription groups is practically the same, only 3 ($\epsilon = 0.9$) up to 11 ($\epsilon = 0.3$) groups remain intact when demand elasticity is accounted for.

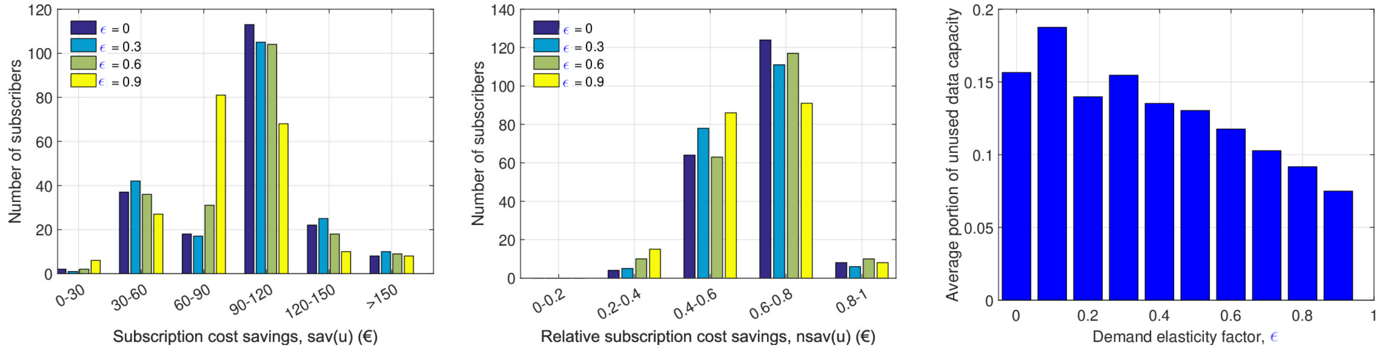
The impact on the data plan assignment is better reflected in Fig. 7c. As users self-control their consumption more aggressively (*i.e.*, higher ϵ), the algorithm tends to assign "tighter" data plans to them since it discounts penalty charges due to consumption beyond the cap. The distribution of per user achievable savings in absolute terms shows a slight shift towards lower values in Fig. 7a. However, this mitigates in relative terms in Fig. 7b. There, we can see that more than 50% of the subscribers (more than 100 out of the 200) reduce their mobile data subscription costs by at least 60% and more than 90% of them (more than 180 out of the 200) save at least 40%, irrespective of how elastic their demand is to the cap and the charged penalty rates upon excess consumption.

VII. CONCLUSIONS

We have considered the generalization of shared capped data plans to *open* groups of users, beyond *e.g.*, the family context or the use of multiple devices by a single user. With open groups the sharing of the fixed fee and overage charges is not straightforward. Hence, we have taken a couple of steps in this paper that close gaps in the relevant literature (ref. section I-A). We have first proposed a cost-sharing rule that is suited to the task. We then proceeded with formulating the joint problem of organizing users into subscription groups and assigning shared data plans to them, as faced by an online platform that issues recommendations about shared data plans to mobile cellular subscribers. This is shown to be an NP-hard problem, hence, as a final step, we devised and extensively evaluated three heuristic clustering-based algorithms that solve it.

Under the ideal assumption that the users' data consumption can be perfectly predicted, all three algorithms achieve subscription savings beyond 50% for at least 70% of users and smaller but still significant savings for the rest of them. As we relax the assumption of perfect prediction and more aggressively underestimate the actual data consumption, the subscription savings are reduced but remain significant: indicatively, even under 10% bias in prediction, the subscription charges are at least halved for 65% of the subscribers when the best of the three algorithms, ACMC, is used. These savings are largely preserved when we explicitly consider that users' data consumption is elastic and adapts to the data plan cap.

In light of this positive experimental evidence about the efficiency of the proposed heuristics, the design of provably efficient algorithms for the problem, possibly leveraging mainstream algorithmic techniques, is the main open technical question. From an economic point of view, what counts as potential savings for the mobile cellular subscribers translates into potential loss for the mobile cellular network operators (MNOs). This appears to be the dominant approach of MNOs to plan sharing so far and explains the rather limited offer in shared plans by MNOs as well as their orientation towards *closed* groups (*e.g.*, family members). Further evidence of this viewpoint is the "penalty" fee charged for each additional user/device added to a shared plan. On the other hand, shared data plans can provide cellular users "with what they want", demand elasticity aside, with less bandwidth and tighter management of network resources. Moreover, they could become a service differentiation factor and evolve to a competitive advantage if they attract interest from certain social groups (*e.g.*, students) or become a hype due to a broader trend for "sharing". It might then become an interesting exercise, on the



(a) Absolute savings (b) Relative savings (c) Plan utilization

Fig. 7. Distribution of per user actual and relative subscription savings and utilization of data plan caps under demand elasticity : $U = 200$.

operator's side, how to design these shared plans' parameters (caps, excess charges) to end up on a positive overall balance without the need of (high) plan-sharing penalties.

Acknowledgment

The research work of M. Karaliopoulos received funding from the Hellenic Foundation for Research and Innovation (HFRI) and the Hellenic General Secretariat for Research and Innovation (GSRI), under grant agreement No 892. I. Koutsopoulos acknowledges the support from the CHIST-ERA grant CHIST-ERA-18-SDCDN-004, through GSRI grant number T11EPA4-00056.

References

[1] Cisco Annual Internet Report (2018–2023). White Paper. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.pdf

[2] Hellenic telecommunications post commission. (2019) 2019 Annual Report. [Online]. Available: https://services.eett.gr/areport2019/en/

[3] Statista. (30 Sept. 2021) Where Data Plans Know No Bounds. [Online]. Available: https://www.statista.com/chart/25886/countries-with-highest-percentage-of-users-with-unlimited-mobile-data-plans/

[4] S. Sen, C. Joe-Wong, and S. Ha, "The economics of shared data plans," in *22nd Workshop on Information Technologies and Systems, WITS 2012*, Jan. 2012, pp. 157–162.

[5] Y. Jin and Z. Pang, "Smart data pricing: To share or not to share?" in *Proc. INFOCOM 2014 WKSHPS*, 2014, pp. 583–588.

[6] J. C. Cardona, R. Stanojevic, and N. Laoutaris, "Collaborative consumption for mobile broadband: A quantitative study," in *Proc. 10th ACM CoNEXT*, New York, NY, USA, 2014, p. 307–318.

[7] Y.-J. Chen and K.-W. Huang, "Pricing data services: Pricing by minutes, by gigs, or by megabytes per second?" *Information Systems Research*, vol. 27, no. 3, p. 596–617, Sep. 2016.

[8] L. Zheng, C. Joe-Wong, M. Andrews, and M. Chiang, "Optimizing data plans: Usage dynamics in mobile data networks," in *Proc. IEEE INFOCOM 2018*, Honolulu, HI, USA, April 2018, pp. 2474–2482.

[9] Z. Wang, L. Gao, and J. Huang, "Multi-dimensional contract design for mobile data plan with time flexibility," in *Proc. 18th ACM Mobihoc*, Los Angeles, CA, USA, 2018, p. 51–60.

[10] X. Wang and L. Duan, "Economic analysis of rollover and shared data plans," *IEEE Transactions on Mobile Computing*, vol. 19, no. 9, pp. 2088–2100, 2020.

[11] Z. Wang, Y. Zhuang, Z. Wang, and X. Wu, "Individual data plan in virtual network operation: A proactive matching approach," in *Proc. IEEE INFOCOM 2019*, 29 April-2 May 2019, pp. 271–279.

[12] R. Venkatesh and W. Kamakura, "Optimal bundling and pricing under a monopoly: Contrasting complements and substitutes from independently valued products," *Journal of Business*, vol. 76, no. 2, p. 211–231, 2003.

[13] L. Weatherford and S. Pölt, "The price of anarchy of serial, average and incremental cost sharing," *Journal of Revenue and Pricing Management*, vol. 3, no. 1, pp. 234–254, October 2002.

[14] H. Moulin and S. Shenker, "Average cost pricing versus serial cost sharing: An axiomatic comparison," *Journal of Economic Theory*, vol. 64, no. 1, pp. 178–201, 1994.

[15] H. Moulin, "The price of anarchy of serial, average and incremental cost sharing," *Economic Theory*, vol. 36, no. 3, pp. 379–405, 2008.

[16] G. Cheirmpos, M. Karaliopoulos, and I. Koutsopoulos, "Optimizing shared data plans for mobile data access," in *Proc. 33th International Teletraffic Congress (ITC-33)*, 2021, pp. 1–9.

[17] F. Miksa, L. Moser, and M. Wyman, "Restricted partitions of finite sets," *Canadian Mathematical Bulletin*, vol. 1, no. 2, pp. 87–96, 1958.

[18] V. Moll, J. Ramirez, and D. Villamizar, "Combinatorial and arithmetical properties of the restricted and associated bell and factorial numbers," *Journal of Combinatorics*, vol. 9, no. 4, pp. 693–720, 2018.

[19] A. K. Chakravarty, J. B. Orlin, and U. G. Rothblum, "A partitioning problem with additive objective with an application to optimal inventory groupings for joint replenishment," *Oper. Res.*, vol. 30, no. 5, Oct 1982.

[20] R. S. Garfinkel and G. L. Nemhauser, "The set-partitioning problem: Set covering with equality constraints," *Operations Research*, vol. 17, no. 5, pp. 848–856, 1969.

[21] F. Kelly, "Notes on effective bandwidths," *In Stochastic Networks: Theory and Applications (Editors F.P. Kelly, S. Zachary and I.B. Ziedins) Royal Statistical Society Lecture Notes Series, 4*, pp. 141–168, 1996.

[22] Ericsson mobility report. (2017) Shifting mobile data consumption and data plans. [Online]. Available: https://www.ericsson.com/en/mobility-report/articles/shifting-mobile-data-consumption-data-plans

[23] F. J. Massey, "The Kolmogorov-Smirnov Test for Goodness of Fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.

Merkouris Karaliopoulos received his Diploma in Electrical and Computer Engineering from the Aristotle University of Thessaloniki, Greece, and a Ph.D. in Electronic Engineering from the University of Surrey, UK. Since 2016 he has been a Senior Researcher at the Athens University of Economics and Business, Greece. Prior to that, he held research/teaching- appointments at the University of Athens (Greece), ETH Zurich (Switzerland), and the University of North Carolina at Chapel Hill (United States). His research interests lie in performance analysis and resource allocation problems emerging in wireless networks and online/mobile platforms.

This article has been accepted for publication in IEEE Transactions on Network and Service Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSM.2022.3231840

Georgios Cheirmpos received the B.Sc. degree in Computer Science from the Athens University of Economics and Business, in Greece, in 2021 and the M.Sc degree in Data Science from the University of Amsterdam, Netherlands, in 2022. He has been with the Research Content Operations team of Elsevier since early 2022 as a Data Scientist working on important pre-production and pilot projects of the team. His research interests include Natural Language Processing, Machine Learning, Data Engineering and Optimization.

Iordanis Koutsopoulos (Senior Member, IEEE) received the Diploma degree in Electrical and Computer Engineering from the National Technical University of Athens (NTUA), Greece, in 1997, and the M.Sc. and Ph.D. degrees in Electrical and Computer Engineering from the University of Maryland, College Park, MD, USA, in 1999 and 2002 respectively. He is currently Professor with the Department of Informatics of Athens University of Economics and Business (AUEB), Greece and Director of the M.Sc. program on Computer Science. He has been with AUEB since 2013 and has served as Assistant Professor (2013-2016) and Associate Professor (2016-2021) in the same department. From 2005 to 2013, he was with the Department of Electrical and Computer Engineering, University of Thessaly, Greece. His research interests include control, optimization and applied machine learning, with applications in mobile computing, wireless networks, social networks, recommendation systems, smart energy grid, and cloud computing. He has co-authored over 170 papers in international journals and conferences. He received the Fulbright Scholarship (1997-2003), the Marie Curie Fellowship (2005-2007), a single-investigator European Research Council (ERC) Competition Runner-Up Award (2012-2015), and three Best Paper Awards for research on online advertising, network economics, and network experimentation