SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK

TECHNICAL REPORT NO. 331

April 1977

INTERNAL TELEPHONE BILLING RATES - A NOVEL APPLICATION OF NON-ATOMIC GAME THEORY

by

Louis J. Billera
David C. Heath
Joseph Raanan

Research supported in part by the National Science Foundation under grants MPS75-02024 and ENG75-00570, and by the Office of Naval Reeearch under contract N00014-75-C-0678.

Abstract

The problem of determining rates is considered for a situation in which services are purchased in bulk, but they have to be paid for by a large number of small users. The desired rates must be "fair" and they must cover all costs. The problem is formulated as a non-atomic game and solved by using the value of the game. In addition to the general problem, a detailed actual case is presented together with computational methods and results.

Introduction

The problem of deciding how to divide the costs of services that are generated and paid for in bulk among many "small" or even "infinitesimal" users arises frequently in many different areas. Examples are the division of overhead costs incurred by a large corporation among its subsidiaries, the assessment of fees or taxes to pay for a municipal service, computer center costs and the determination of the internal billing rates for long-distance telephone calls that are placed through WATS (Wide Area Telecommunication Service) lines.

We show how a typical one of these problems can be formulated as non-atomic games. We then show how, among the various solution concepts for the games the Aumann-Shapley value (see Aumann-Shapley [1]) is the one that best answers the problem, given that the desired rates must satisfy certain reasonable requirements (including fairness, efficiency, etc.).

We present the general problem in Section 1. Section 2 presents the particular problem of determining internal telephone billing rates. Section 3 summarizes the concepts of non-atomic game theory that are needed, and Section 4 makes the connection between the problem and the model. Computational methods and results are presented in Section 5.

1. The General Rates Problem for Service in Bulk.

The general rates problem for services that are available in bulk can be described as follows: There is a group of customers, or consumers requiring some service. The service cannot be provided to each customer separately in exactly the amount he needs, but rather is available in bulk. By that we mean that it can either be supplied in big "chunks" or that there is a very high initial fee involved that makes it unprofitable for any single consumer to purchase that service by himself. However, if the customers form a coalition, their total demand may be high enough to make acquiring that service feasible. The question that arises now is how to charge each customer his "fair share" of the costs, given that we have to recover all of it from users' payments. The problem becomes more involved if there is more than one type of service, or if there are different grades of service. In addition, there may be some "overhead" charges involved that must be paid for by the users. Usually, the overhead, too, cannot be broken down in such a way as to enable us to determine what part of it was directly caused by each consumer. Another difficulty may be presented by the fact that the service required by each customer is "infinitesimal" and that only a very large group of customers could profit by purchasing the service.

These problems fall naturally into the realm of cooperative game theory. There, too, we have players (i.e., customers as defined by their service requirements) who can form coalitions and gain by doing so; different coalitions may gain more than others. The question there, too, is how to divide the "benefits of cooperation" among the players. Depending on the criteria we want our solution to satisfy, we get different solutions. Among these, there is one that, as will be shown here, is particularly suitable for the problem at hand, namely, the value of the game (see Shapley [5] and Aumann-Shapley [1]). This gives us the unique set of rates that are fair (in a sense that will be made precise in Section 3), efficient (i.e., total revenue = total cost) additive (over coalitions) and linear (over games).

2. The Problem of Internal Telephone Billing Rates.

Long-distance telephone service for users in the continental U.S. can be obtained in the following ways:

A. Direct Distance Dialing (DDD), where the call is charged, basically, by the length of the call and the distance to its destination.

B. Foreign Exchange (FX) lines, where the user has a telphone connected to an exchange in another area. The charge consists of a monthly fee for the line rental (related to the distance) plus the regular cost of a telephone at that foreign exchange.

C. Wide Area Telecommunications Service (WATS). The area of the continental U.S. outside the user's state is divided into 5 concentric bands. Band $i+1$ contains band i, $i = 1,2,3,4$. Band 1 usually contains just the neighboring states, while band 5 contains all of the continental U.S. WATS line servicing those bands can be obtained from the phone company under two different plans. Under both plans the user pays for the time the lines are used, regardless of the destination of the calls; both plans consist of an initial fee - covering an initial amount of time - and an incremental charge for each hour above the inital allocation. The difference between the two plans is that one offers a very large initial amount of time (which is rarely fully used, let alone exceeded) and a small charge for incremental use, and the other offers a limited initial time but a higher incremental charge.

The following questions arise:

(i) Which, if any, of these services should be bought?

(ii) If any services are bought, how to allocate the charges to the users.

Question (i) was addressed by Heath, Lampell and Prabhu in [2]. There, a queueing model was built of a telephone system and a computer code was written that generates the optimal (least-cost) line configuration for a given demand, using a Branch and Bound algorithm. If the demand is, indeed, not high enough to justify any WATS or FX lines, the solution will be the 0 configuration.

The telephone system considered included a computerized device that can route each call coming into it onto the best sequence of WATS or FX lines. For example, if the incoming call has an area code that is within WATS band 3, the device will try to place it on a WATS 3 line. Failing that WATS 4 and WATS 5 will be tried, and finally - if all that fails due to unavailability of free lines - the call will be sent DDD. The device also provides security checks to prevent unauthorized users from placing a call and a detailed bookkeeping information - who called, where to, how long the call lasted, etc.

Notice, however, that this further complicates question (ii), since the charges involved with the device are mainly fixed - rental, maintenance, operator's salary - and are not directly associated with the calls. Hence, these costs, too, cannot be divided among the calls in any straightforward manner.

At first, it seems that some of the cost incurred can be directly attributed to the individual calls. For example, message units for calls using FX lines or the incremental charges for calls that cause these charges are such costs. However, for the former the question of allocating overhead expenses remains unsolved while for the latter, the question is even more complicated since the timing of calls accumulates on a monmtly basis, the first few calls to use the line will utilize the initial time included in the initial fee and will not incur any incremental charges, and all the incremental charges will be incurred by calls placed later in the month. But, is it indeed right to charge those calls differently? The answer seems to be no.

The client imposed the following constraints on the rates:

A. Since the client was mandated to break even in this operation, the rates must exactly cover the expenses incurred in providing the telephone services.

B. The rates must be "fair", or "symmetric" since calls are charged to different accounts and budgets (such as Research Grants, administrative funds, etc.), and all must be charged the same rates. That is, in short, two calls made to the same destination during the same period in the day must be charged the same rate regardless of their purpose, the account they will be charged to or the person or office that placed them.

3. Non-Atomic Games and Their Values.

The subject of non-atomic games and their values was extensively studied by Aumann and Shapley [1]. We will introduce here the main notions and results of this theory that are relevant to our discussion here. All definitions are as they appear in [1].

Let $(I,\mathcal{C})$ be a measurable space, where I is a set and $\mathcal{C}$ is a

σ-field of measurable subsets of I. The members of I are called players and the members of $\mathcal{C}$ coalitions. A game v on I is a real-valued set function on $\mathcal{C}$ satisfying $v(\phi) = 0$. A game v is monotonic if for all $S,T \in \mathcal{C}$, $S \subset T \Rightarrow v(S) \leq v(T)$. A game v is of bounded variation if it is the difference between two monotonic games. The space of all games of bounded variation is called BV.

If Q is any space of games, Q^+ denotes the cone of the monotonic members of Q. A mapping of Q into BV is called positive if it maps Q^+ into BV^+.

The subspace of BV consisting of all bounded, finitely additive set functions (i.e., the bounded, finitely additive signed measures on $(I,\mathcal{C})$) is denoted FA. Denote by NA the space of non-atomic measures on $(I,\mathcal{C})$. (A measure μ is non-atomic if for all $S \in \mathcal{C}$, $|\mu(S)| > 0 \Rightarrow \exists\, T \subset S$, $T \in \mathcal{C}$, with $0 < |\mu(T)| < |\mu(S)|$). Clearly $NA \subset BV$. The subspace of BV spanned by all powers of NA^+ measures will be denoted by pNA.

Let $\mathcal{G}$ be the group of automorphisms of $(I,\mathcal{C})$. (I.e., the one-to-one mappings of I onto itself that are measurable in both directions). Each $\theta \in \mathcal{G}$ induces a linear mapping θ_* of BV onto itself defined by $(\theta_* v)(S) = v(\theta S)$ for all $S \in \mathcal{C}$. A subspace Q of BV is called symmetric if $\theta_* Q = Q$ for all $\theta \in \mathcal{G}$.

We now come to the definition of value. Let Q be a symmetric subspace of BV. A value on Q is a positive linear mapping ϕ from Q into FA that satisfies the following axioms:

A1. The efficiency axiom: $(\phi v)(I) = v(I)\ \forall\, v \in Q$.

A2. The symmetry axiom. $\phi\theta_* = \theta_*\phi\ \forall\, \theta \in \mathcal{G}$.

The major result which concerns us here is given by the following theorem ([1], p. 23, theorem B):

There is a unique value ϕ on pNA. Furthermore, let μ be a vector of measures in NA, and let f be continuously differentiable on the range R of μ with $f(0) = 0$. Then $f\circ\mu \in$ pNA and, when R has full dimension,

$$(3.1)\ \phi(f\circ\mu)(S) = \sum_{j=1}^{n} \mu_j(S) \int_0^1 f_j(t\mu(I))dt,$$

where f_j denotes $\partial f/\partial x_j$.

4. The Telephone Game.

We now show how the theory of non-atomic games enables us to solve the rates problem. We have a monthly collection of calls which we denote by I (and which is the underlying space). Each calling instant is thus a "player". We now break down this collection according to the time of day during which the calls were placed, their destination and the type of day on which they were made (business day - called High-Use Day - or weekend). If we have k different destinations (number of WATS bands and FX lines, etc., in the system), we then have $n = 24 \times k \times 2$ different "types" of calls. We now define n measures on I, μ_j, $j=1,\ldots,n$, where for any subset of calls $S \subset I$ $\mu_j(S)$ is the total number of minutes of telephone calls of type j in the subset S. For example, $\mu_{23}(S)$ may measure the total number of minutes of telephone calls in S that were placed during the month to WATS band 3, between 2 and 3 a.m., on a business day. We let $\mu = (\mu_1,\ldots,\mu_n)$ be the vector of the measures.

Given a certain load X on the system, $X = (x_1,\ldots,x_n)$ we can use the optimization routine (see [2]) to find the least-cost configuration to service that load. We denote the cost of this configuration (including any "overhead") by $f(x_1,\ldots,x_n)$. Now we have a game v on I, defined by $v = f\circ\mu$ that is, for each $S \subset I$ we have $v(S) = f(\mu(S))$, which is the minimal cost of servicing the demand for phone calls represented by S.

We now want a solution to this game that will enable us to determine the right price to charge for each minute of phone call of each of our n types. We note that, indeed, the Shapley value is the right solution. The efficiency axiom A1 is exactly the requirement that total revenue equal total costs; the symmetry axiom is the "fairness" condition.

The value is unique if we agree to add to these requirements the linearity condition. This condition means here that if we could break the costs into their various components such as line rental, maintenance, taxes, employees' salaries, etc. and then assign each minute of phone call its "fair share" of these costs, then these shares would add up to exactly the charge for that minute of phone call if we combined all these expenses together at the outset. This seems a very reasonable assumption to us, and it was indeed introduced.

Let us now examine carefully the formula for the value, as given by (3.1). We have for any $S \subset I$ that

$$\phi(f\circ\mu)(S) = \sum_{j=1}^{n} \mu_j(S) \int_0^1 f_j(t\mu(I))dt$$

Now, $\phi(f\circ\mu)(S)$ represents the part of the total cost of the system that must be borne by S. We notice that if S consists only of all the calls of one type, say type k, then we will have that

$$\mu_j(S) = 0 \quad \text{unless} \quad j = k.$$

so, $$\phi(f\circ\mu)(S) = \mu_K(S) \int_0^1 f_K(t\mu(I))dt,$$

and this is the cost that must be borne by all the calls of type k. Since $\mu_k(S)$ is the total number of minutes of phone calls of type k in the month,

$$r_k = \int_0^1 f_k(t\mu(I))dt$$

is the rate, per minute, that must be charged to those calls; and for any $S \subset I$, we see from (3.1) that the value of this coalition S - which here means its share of the costs - is exactly the sum over $j = 1,\ldots,n$, of the amount of calls of type j in that coalition of calls times r_j, so the r_j's are indeed rates. It was these r_j's that were computed and proposed as the right rates to be charged to the various types of calls.

Notice that these rates depend on the behaviour of the function f on the diagonal $\{t\mu(I) | 0 \leq t \leq 1\}$, i.e., on the cost of servicing a uniformly shrinking load. This means that r_k is exactly the average of the marginal costs for calls of type k when the system is enlarged uniformly from 0 to $\mu(I)$.

Note: The theorem quoted above, and formula (3.1), were stated for a function f satisfying certain regularity conditions. Our cost function, however, does not satisfy these regularity conditions precisely. So the theory had to be extended to cover that case. This was done and will appear elsewhere (see Raanan [4]).

5. The Rates.

During the first phase of the project, a number of computer codes were written [2]. Among those were programs that were designed to perform the following tasks:

A. For a given monthly load, calculate the optimal WATS and FX line configuration. This program will be referred to as the optimizing routine.

B. Given a configuration of WATS and FX lines and a monthly load, compute the cost of servicing that load using the given configuration. This program will be referred to as the cost-calculation routine.

We recall that we needed to compute r_j, $j = 1,\ldots,n$ where

$$r_j = \int_0^1 f_j(t\mu(I))dt$$

where $f_j = \partial f/\partial x_j$. We know that when $t = 0$, no WATS or FX lines are needed. We also found out that when $t = 1$, it was optimal to have a number of WATS and FX lines. Define t_0 as follows:

$$t_0 = \min\{t \mid 0 < t \leq 1 \text{ and for the load } t\mu(I) \text{ it is optimal to have at least 1 WATS or FX line together with the computerized device}\}.$$

We can now rewrite r_j:

$$r_j = \int_0^{t_0} f_j(t\mu(I))dt + \int_{t_0}^1 f_j(t\mu(I))dt$$

The reason for this breakdown is that for $0 \leq t \leq t_0$ there are no WATS or FX lines so all calls are sent DDD. Since f_j is the marginal cost of a call of type j, $f_j(t\mu(I))$ will be equal to the DDD cost for calls of type j for $0 \leq t \leq t_0$.

So, the first step was to calculate t_0. For this purpose, the "overhead" costs were added to the costs of the optimal configuration for a load of $t\mu(I)$ and this was compared with the DDD cost of that load, for various values of t. Once t_0 was known, the process continued, as follows:

t was varied from t_0 to 1 in fixed increments Δt. For each such t the optimal configuration for a load of $t\mu(I)$ was calculated. Next, a modified cost-calculating routine produced the numerical partial derivatives and these were integrated from t_0 to 1. Then, the DDD rates were multiplied by t_0 and added to the integrals, thus producing the rates as shown in Table I. These rates, however have the undesired property that, although they are the right rates to use, they are different for almost every hour during the working hours. To remedy this, weighted averages of the rates r_j -- weighted by the loads μ_j - were calculated. We noticed that the usage - and the rates - were very close for the time periods between 8 and 9 a.m. and between 12 noon and 1 p.m., so those were averaged. The rest of the business-day rates were close, so the rates for the periods of between 9 a.m. and 12 noon and 2 to 5 p.m. were averaged. For the remaining period, that between 1 and 2 p.m., the rate was taken to be the arithmetic average of the 12 to 1 p.m. rate and the 2 to 5 p.m. rate following the rates in Table I. These new rates are presented in Table II. For comparison, Table III gives the DDD rates for the hours of the business day (8 a.m. to 5 p.m.) for the various WATS bands and FX lines.

TABLE I

TABLE OF RATES BY BAND AND TIME OF DAY IN CENTS PER MINUTE
HIGH-USE DAY
PRECISE-UNAVERAGED - RATES

TIME	BAND WATS 1	D.C.	WATS 2	WATS 3	WATS 4	WATS 5	N.Y.C.	ALBANY	SYRACUSE	BUFFALO	WATS 0	AC607
MDNT-100	7.9	8.4	8.1	8.5	8.8	9.6	11.9	10.6	8.7	11.2	12.9	9.0
100-200	7.9	8.4	8.1	8.5	8.8	9.5	11.9	10.6	8.7	11.2	12.9	9.0
200-300	7.9	8.4	8.1	8.5	8.8	9.5	11.9	10.6	8.7	11.2	12.9	9.0
300-400	7.9	8.4	8.1	8.5	8.8	9.5	11.9	10.6	8.7	11.2	12.9	9.0
400-500	7.9	8.4	8.1	8.5	8.8	9.5	11.9	10.6	8.7	11.2	12.9	9.0
500-600	7.9	8.4	8.1	8.5	8.8	9.5	11.9	10.6	8.7	11.2	12.9	9.0
600-700	7.9	8.4	8.1	8.5	8.8	9.5	11.9	10.6	8.7	11.2	12.9	9.0
700-800	7.9	8.4	8.1	8.5	8.8	9.6	11.9	10.6	8.7	11.3	12.9	9.0
800-900	19.8	20.3	20.5	21.3	22.2	23.9	30.3	28.2	19.5	27.3	29.5	19.4
900-1000	24.5	23.0	26.3	27.1	28.0	30.4	33.5	34.1	22.3	31.2	36.4	24.5
1000-1100	26.3	24.1	30.4	31.3	32.1	34.5	35.1	34.5	23.1	33.0	36.7	24.7
1100-1200	29.2	25.2	31.7	32.6	33.5	35.9	36.0	33.2	21.4	33.5	36.7	24.7
1200-1300	21.9	20.8	24.5	25.3	26.2	29.1	30.4	27.6	19.1	28.2	28.3	18.5
1300-1400	26.7	23.3	29.0	29.9	30.8	33.5	32.0	31.3	21.0	31.5	34.0	23.0
1400-1500	30.8	26.8	33.1	33.9	34.9	37.3	35.5	34.2	22.6	32.6	36.8	24.8
1500-1600	30.3	26.1	32.5	33.4	34.3	36.7	34.8	32.8	21.9	32.0	35.9	24.1
1600-1700	26.9	23.5	29.8	30.6	31.5	34.2	31.1	29.3	19.7	29.9	31.1	20.7
1700-1800	13.2	13.4	14.4	15.0	15.6	17.7	15.2	14.9	14.2	15.4	16.5	13.0
1800-1900	12.8	13.4	13.3	13.8	14.4	15.7	15.2	14.5	14.2	15.2	17.9	12.5
1900-2000	12.8	13.4	13.3	13.8	14.4	15.5	15.2	14.5	14.2	15.2	17.9	12.5
2000-2100	12.8	13.4	13.3	13.8	14.4	15.5	15.2	14.5	14.2	15.3	18.0	12.5
2100-2200	12.9	13.4	13.3	13.9	14.4	15.6	15.2	14.5	14.2	15.3	18.0	12.5
2300-MDNT	7.9	8.4	8.2	8.5	8.9	9.6	11.9	10.6	8.7	11.2	12.9	9.0

TABLE II

TABLE OF RATES BY BAND AND TIME OF DAY IN CENTS PER MINUTE
HIGH-USE DAY

AVERAGED RATES

BAND

TIME	WATS 1	D.C.	WATS 2	WATS 3	WATS 4	WATS 5	N.Y.C.	ALBANY	SYRACUSE	BUFFALO	WATS 0	AC607
MDNT-100	7.9	8.4	8.1	8.5	8.8	9.6	11.9	10.6	8.7	11.2	12.9	9.0
100-200	7.9	8.4	8.1	8.5	8.8	9.5	11.9	10.6	8.7	11.2	12.9	9.0
200-300	7.9	8.4	8.1	8.5	8.8	9.5	11.9	10.6	8.7	11.2	12.9	9.0
300-400	7.9	8.4	8.1	8.5	8.8	9.5	11.9	10.6	8.7	11.2	12.9	9.0
400-500	7.9	8.4	8.1	8.5	8.8	9.5	11.9	10.6	8.7	11.2	12.9	9.0
500-600	7.9	8.4	8.1	8.5	8.8	9.5	11.9	10.6	8.7	11.2	12.9	9.0
600-700	7.9	8.4	8.1	8.5	8.8	9.5	11.9	10.6	8.7	11.2	12.9	9.0
700-800	7.9	8.4	8.1	8.5	8.8	9.6	11.9	10.6	8.7	11.3	12.9	9.0
800-900	21.1	20.6	23.1	24.3	24.8	28.7	30.4	28.0	19.4	28.1	29.0	19.0
900-1000	28.7	24.9	30.9	32.1	32.8	35.7	34.7	33.4	22.1	32.2	36.0	24.2
1000-1100	28.7	24.9	30.9	32.1	32.8	35.7	34.7	33.4	22.1	32.2	36.0	24.2
1100-1200	28.7	24.9	30.9	32.1	32.8	35.7	34.7	33.4	22.1	32.2	36.0	24.2
1200-1300	21.1	20.6	23.1	24.3	24.8	28.7	30.4	28.0	19.4	28.1	29.0	19.0
1300-1400	24.9	22.8	27.0	28.2	28.8	32.2	32.5	30.7	20.8	30.2	32.5	21.6
1400-1500	28.7	24.9	30.9	32.1	32.8	35.7	34.7	33.4	22.1	32.2	36.0	24.2
1500-1600	28.7	24.9	30.9	32.1	32.8	35.7	34.7	33.4	22.1	32.2	36.0	24.2
1600-1700	28.7	24.9	30.9	32.1	32.8	35.7	34.7	33.4	22.1	32.2	36.0	24.2
1700-1800	13.2	13.4	14.4	15.0	15.6	17.7	15.2	14.9	14.2	15.4	18.5	13.0
1800-1900	12.8	13.4	13.3	13.8	14.4	15.7	15.2	14.5	14.2	15.2	17.9	12.5
1900-2000	12.8	13.4	13.3	13.8	14.4	15.5	15.2	14.5	14.2	15.2	17.9	12.5
2000-2100	12.8	13.4	13.3	13.8	14.4	15.5	15.2	14.5	14.2	15.3	18.0	12.5
2100-2200	12.9	13.4	13.3	13.9	14.4	15.6	15.2	14.5	14.2	15.3	18.0	12.5
2300-MDNT	7.9	8.4	8.2	8.5	8.9	9.6	11.9	10.6	8.7	11.2	12.9	9.0

TABLE III

DDD rates to the various zones on a weekday, 8 a.m. - 5. p.m.*

WATS Band	W1						
or Fx	W1	W2	W3	W4	W5	WO	Washington, D.C.
Rate (¢/minute)	35.4	36.6	38.0	39.5	42.5	42.0	35.7
	New York City	Albany	Syracuse	Buffalo	Area Code 607		
	52.8	48.0	33.9	47.6	24.7		

* The rates have been adjusted to match the timing provided by the computerized device.

6. Summary.

We have presented here a new method for determining equitable rates for a large class of problems using the concept of the value of an associated game. While the value of a game has been used to determine rates in a situation involving finitely many players and a special cost structure (see Littlechild and Owen [3]), the present work represents (to our knowledge) the first application of the value of a non-atomic game for this purpose. In the problem considered here, the non-atomic nature of the game is inherent; the "players" represent instants of telephone calls for which a continuous model is most appropriate.

The methods presented here should be widely applicable to similar problems such as (time of day) pricing of utilities such as electricity, steam, water, and so on. In such problems the service is provided in continuously variable amounts and the costs involve both large fixed costs and non-linear variable costs.

The rates developed by this procedure have, with minor modifications, been adopted for use at Cornell University.

Acknowledgements

We are grateful to the Department of Planning and Facilities at Cornell University for proposing the problem and for partially supporting this research.

References.

[1] Aumann, Robert J. and Lloyd S. Shapley, Values of Non-Atomic Games, Princeton University Press, Princeton, N.J. 1974.

[2] Heath, David C., David M. Lampell and N.U. Prabhu, Optimization of Toll-Calling Facilities, Technical Report No. 330, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, N.Y., April 1977.

[3] Littlechild, S.C. and G. Owen, A Simple Expression for the Shapley Value in a Special Case, Management Science, Vol. 20, November, 1973, pp. 370-372.

[4] Raanan, Joseph, Ph.D. dissertation, Cornell University, Ithaca, New York, in preparation.

[5] Shapley, Lloyd S., A Value for n-Person Games, pp. 307-317 in Contributions to the Theory of Games II, Mathematics Studies Number 28, Princeton University Press, Princeton, N.J., 1953.