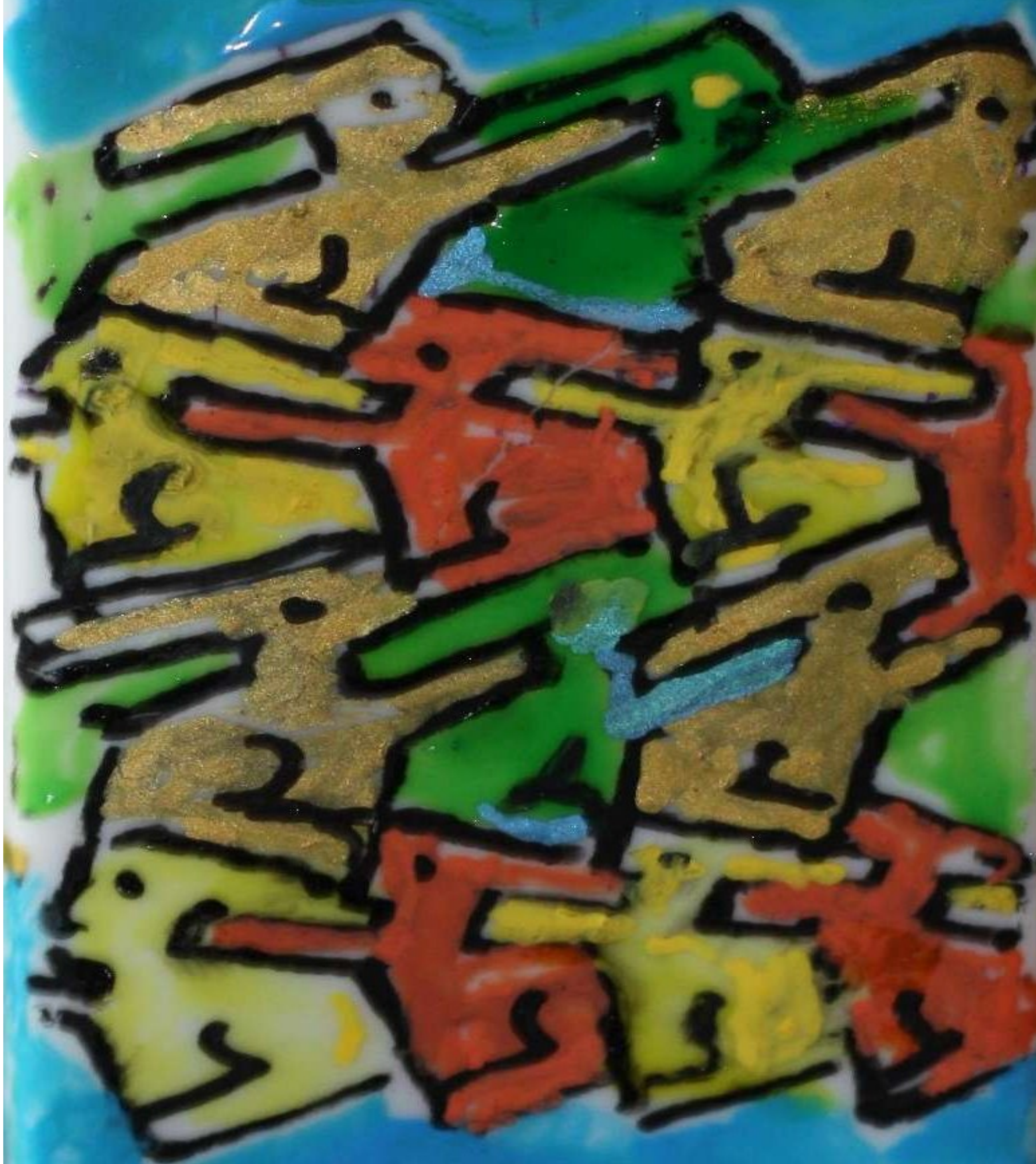
NETWORKING GAMES



EITAN ALTMAN and CORINNE TOUATI

Networking Games

Eitan Altman[1] and Corinne Touati[2]

[1]INRIA Méditerranée, France, Email: altman@sophia.inria.fr
[2]INRIA Rhône Alpes, France, Email: corinne.touati@inria.fr

Todo list

Contents

Chapter 1

Introduction

1.1 Some history

1.1.1 The Talmud Example

The oldest reference to game theory that appears in references on the history of this disipline goes back to the TALMUD (200 AD) which is a Jewish theological book, which serves also as the basis for the jewish religious law. The following situation appears there and serves to explaian how to split costs among players. A man dies leaving an estate of size e and debtbs of size d1,...,dn that exceeds e. How much should each creditor get? The Talmud answers by providing the following three examples.

- If e= 100 then d=(33.3, 33.3, 33.3)
- If e= 300 then d=(50, 100, 150),
- If e= 200 then d=(50, 75, 75).

The question of how the answers were generated in each of the examples remained an open question. It was solved less than 30 ago by Aumann and Maschler. The authors published their finding solution simultaneously in a mathematical and in a theological journal.

The explanation is based on other rules that handle the simpler case of splitting between two persons. The general question is of how to split some value when the sum of claims of the two persons exceeds the available amount. To formulate the rule, the Talmud gives two examples.

- Rule 1: If 2 persons have the same claim they receive equal shares
- Rule 2: Assume there is a dispute over some asset, and that player 1 claims that all the asset is due to him, while the second has a claim over half of the asset. Then they split the part that is under dispute (one hald of the asset) equally between the persons, leaving the whole undisputed part to the first player.

Going back to the case of more than two persons, Aumann and Maschler show that there is a unique division between all players such that the following holds. Choose any two players, along with the amount that they receive together. Then the amount received by each of these players is the same as the one that they would obtain if we applied Rule 2 to the problem of splitting that amount between them.

1.1.2 The conflict between Abram and Lot

The story of Abraham and Lot's conflict [Bible, Book of Genesis] describes the separation of the brothers as a result of a fight among their shepherds. The dispute ends with a pieceful agreement which is described in Wikipedia as being the most ancient peace accord known to date. Abraham offers Lot to separate, in order to prevent the fight, and he grants Lot with the right to be the first among the two to pick the territory he desires:

"5 And also Lot, who went with Abram, had flocks and cattle and tents."

"6 And the land did not bear them to dwell together, for their possessions were many, and they could not dwell together."

"7 And there was a quarrel between the herdsmen of Abram's cattle and between the herdsmen of Lot's cattle, and the Canaanites and the Perizzites were then dwelling in the land."

"8 And Abram said to Lot, "Please let there be no notions of quarth, for we are brethren."

"9 Is not all the land before you? Please part from me; if [you go] left, I will go right, and if [you go] right, I will go left. "

Lot accepts the peace deal, for the Partition of the Land, and chooses the area of the plain of the Jordan in Sodom area, and the story ends with Abraham and Lot separately settling in different areas of the Land..."

The conflict between Abraham and Lot can be described as a Coordination Game. The peace solution offered by Abraham is a Stackelberg equilibrium. We shall describe both the coordination games as well as of the Stackelberg equilibrium later.

The description is taken from
`http://en.wikipedia.org/wiki/Abraham_and_Lot%27s_Conflict`

1.2 Impact of game theory on publications in networking

We searched in Google scholar the number of documents found when searching for the following types of games: (1) Power control, (2) Flow control, (3) Rate control, (4) Access control, (5) Jamming and (6) Routing games. We searched for documents containing all of the words "XXX wireless networks GT" where GT stands for game theory and where XXX stands for one of games (1)-(6) above. All figures correspond to the date of October 15th, 2010. The number of the occurrences corresponding to each of the cases are summarized in the first row of Table 12.1. The second row of the table was obtained by repeating the above without the word "games theory" (i.e. searching for all the words "XXX wireless networks"). The last row in the matrix was obtained by searching for the words "XXX networks, game theory".

	Power control	Flow control	Rate control	Access control	Jamming	Routing
Wirelesss Networks, GT	27600	19400	25600	27400	3520	16100
Wirelesss Networks	555000	303000	59000	609000	19000	342000
Networks, GT	391000	174000	264000	298000	17000	38000

Table 1.1: The number of citations

If we consider wireless networks, then in each type of gae, the number of documents that fall within game theory forms around than in 5%. The game that was the most studied in the wireless context is the power control one. (This is also the case if we do not specify to wireless, but then it may cover also games that arise in supplying electricity).

Using the software "publish or perish", we find in 2009, 189 documents with all the words "jamming wireless networks" in google, of which 39 further contain "game" and 22 contain "game theory". In 2008 there are 123 documents containing "jamming wireless networks" and only six containing further "game theory". Thus the ratio of papers that use game theory for studying jammig seems to be around 5%.

Next we consider Figure 1.1 that shows how the number of documents that contain all of the words "Routing games Nash equilibrium game theory" vary as a function of the year.

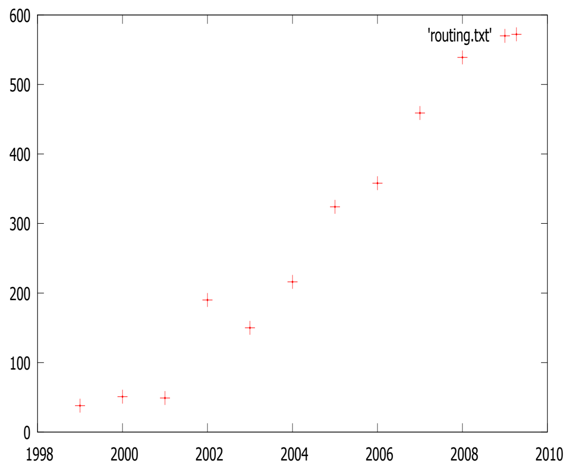


Figure 1.1: The number of documents each year containing all the words "Routing games Nash equilibrium game theory" as a function of the year.

The precise numbers observed each year by searching with scholar google on October 15th, 2010, are given in Table **??**. The total number of documents published on 1999 and earlier has been found to be 217. There is a steady linear growth. Within 10 years, this number is seen to have increased by more than 10.

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
No. of docs	38	51	49	190	150	216	324	358	459	539	570

Table 1.2: The evolution of the number of documents on routing games

Next we use Publish or perish, under cathegory Engineering, Computer Science and Mathematicis and check the popularity of the set of words "game theory Nash equilibrium". We make no particular referene to networking applications. Within the documents published by the IEEE on 2009, 764 were found in 2009 containing the words "Wireless networks". In 1999 in contrasnt we find only 2.

We conclude that there is a significative increase in the impact of game theory as we observe it over documents on the Internet.

1.3 Is game theory an appropriate tool for designing networks

The word "game" may have conotations to "toys" or of "playing" (as opposed to decision making). But in fact it stands for decision making by several decision makers, each having her (or his) own individual objectives. In the special case that there is a common objectives that all players maximize, this is called a team problem. When there is only a single decision maker we speak of optimization rather than team. In a team problem we search for a maximizer of the common objective. In a non-cooperative we search for a typically for a solution at which each player is at a (local) maximum - it cannot do better by a unilateral deviation. This is called an equilibrium.

Remark 1. *Is a game where all players have a common objective to maximize, equivalent to a team problem? Is the equilibrium of the game the same as the solution of the team problem? The answer is no. Any solution of the team problem is an equilibrium to the game problem but the* *converse need not hold. As an example, let there be two players, where player i has to choose either $x_i = 0$ or $x_i = 1$. Consider Then $x_1 = x_2 = 1$ is the team optimal solution. It is also an equilibrium. On the other hand, $x_1 = x_2 = 0$ is an equilibrium but is not an optimal team solution.*

When should one use a game theoretic framework and when should one use a team framework?

Game theory searches for stable solutions to this problem. Is a user of a mobile phone indeed constantly trying to improve his performance? Experience shows that users tend to be coopoerative. In fact, many Internet protocols are very cooperative: TCP that controls the rate of transmission of packets, is an example for a cooperative behavior.

When are the users cooperative? The power control game seems to be the most studied one in wireless networking (according to the figures that we extracted from google). Yet in practice the user does not have access to control the power. The equipment obliges us to be cooperative.

In Section 1.5 we present an example where, in contrast, the equipment provider leaves the decision making to the user and even provides the user the appropriate tools to make and take the decisions.

1.4 Business models of jammers

According to [1], The US military routinely uses jammers to protect secure military areas from electronic surveillance. Jammers can also be used to protect traveling convoys from cell phone triggered roadside bombs in places like Iraq.

Interestingly, the business model of jammer phones include jamming one's own telephone. This allows one to avoid being disturbed. Typical prices of a jammer vary between 100 and 300 USA $.

Leading electronic companies have introduced cellular phone jammers based on the denial of service approach: they simply create noise which interferes with the communications [2]. More efficient techniques have been designed later [2].

Jammers are actually manufactured and sold over the Internet by several companies. Selling jammers in the USA and in Europe is not legal, but it generally is legal in Asia [1]. According to [1], the FCC (Federal Communications Commission) in the United States has outlawed the sale and use of jammers because they can in theory interfere with emergency communications between police and rescue personnel, aid in criminal activity as well as disrupt medical equipment like pacemakers; Using a cell phone jammer may result in fines of upto $ 11,000.

Figure 1.2: Small jammer

Figure 1.3: Big jammer

Figures 1.2 and 1.3 show a a small and a big jammer, respectively. The small one is of the size of an average wireless telephone terminal (photos taken from antennasystems.com and from Globalgadgetuk.com respectively). Most jammers only have a range of about 50 to 80 feet and will only effectively jam their immediate surroundings. Stronger jammers can cover larger structures like office buildings, are also sold. Examples of sites that sell jammers are
www.methodshop.com/gadgets/reviews/celljammers/index.shtml
and www.covert-supply.co.nz/products/Wifi,-Bluetooth,-Wireless-Video-Jammer-%252d-Portable-Wireless-Blocker.html

Most cell phone jammers come in 2 versions, one for Europe, North Africa and the Gulf states GSM networks (900 & 1800) and one for the Americas & Canada (800 & 1900 mhz) networks [1].

References:

Surveys on networking games: In [3], the authors present a survey that focuses on networking games with a special emphasis on the relation between games arising in road traffic networks and those in telecommunication networks. A more general overview on networking games can be found in [4][1]. For a survey focusing on wireless networks, see [5]. A recent survey in French can be found in [6].

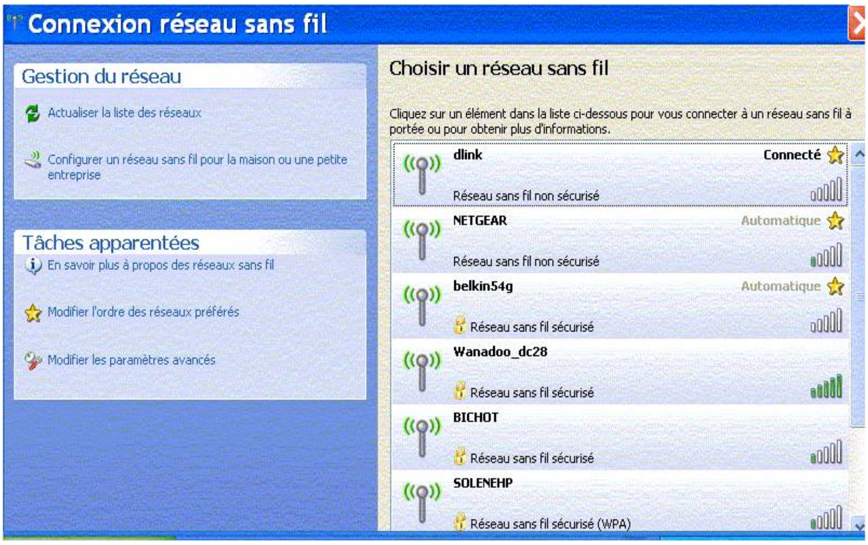


Figure 1.4: The access point association problem: Information available when taking a decision

1.5 The Association problem

There are several types of association games that one is frequently faced with.

Choosing an access point

When atempting to connect to the Internet, one may have the option of choosing between several access points that use wireless local area networks (WLANS). The decision, of which access point to connect to, is typically left to the user, The driver for the wireless card typically gives some information concerning the channel state at each one of the access points. Fig 1.4 is an example of the information presented for a user when the opportunity of taking a decision is offered to him. It is easily seen that the user is indeed put in a situation of a game.

This is a complex stochastic game as each user comes at random points, its decision will be affected by the state of the channel not only at the present (i.e. the one it has available) but also at the future, and the latter will be determined by the decisions of future users and a user is not aware of when future arrival will occur and what the decisions will be.

This game has an unusual information: it is partial and **missleading**. Missleading - because, although the the channel state indeed can gives information on the transmission rate, it is known that the actual throughput of a user is a function of not only his channel state but also of that of the other connected users. (The throughput is known to be lower bounded by the harmonic

[1]see a more recent update in
http://www-sop.inria.fr/members/Eitan.Altman/PAPERS/srvUpdate.pdf

means of the rates available to each user). The real utility of a user is the throughput he would get and the user may not be aware that it is possible that an access point with a better channel may have a lower throughput because more terminals are connected to it.

Choosing between technologies

We may have to choose between several technologies: say between 3G, WIFI, bluetooth and Ad-Hoc. Figure 1.5 is an example of the iniformation available to a user in a game where one has the option of connecting to an Ad-hoc network or to an access point.

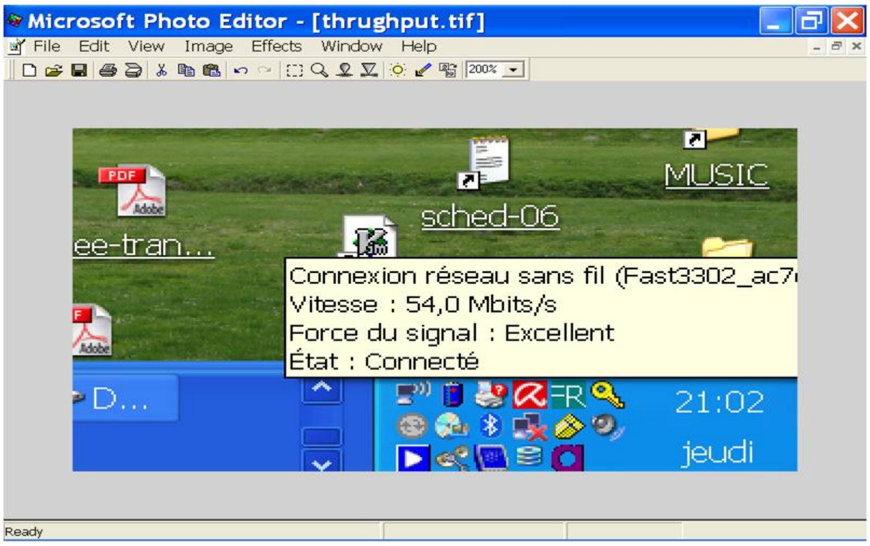


Figure 1.5: The association problem of choice of technology: Information available when taking a decision

Refined modeling of the above games

So far we considered a game where the each user takes one decision: where to connect upon arrival. However, once the user is connected he may get more information about his throughput. An example of the grafical form that such extra information is presented to us is given in Figure **??**. This information too may be missleading. The one we see in Figure **??** is the physical channel rate. Again, the throughpuut of the user is not this channel rate but some function of the channel rates of all usuers; this function is bounded by the harmonic mean of the channel rates of all users connected to the access point.

The new available information could be used to reconsider the connection decision. In that case the game becomes more complex as the strategy of the users is a more complex object.

A hierarchical game of association

In some cases, each of the access points corresponds to that of another operator. In other cases, the choice of operator is offered to a user only once it cnnects to an access point. This is again a game. An example of the way that the choices are presented to the user is presented in Figure **??**.

This game may also be a hierarchical one. It may involve a preliminary decision to which service provider to attempt connection. Once an attempt is made then the user gets an information on the pricing policy of the provider (note that the user may know the pricing in formation of one or more providers before making the decision since this pricing usually remains the same for a long

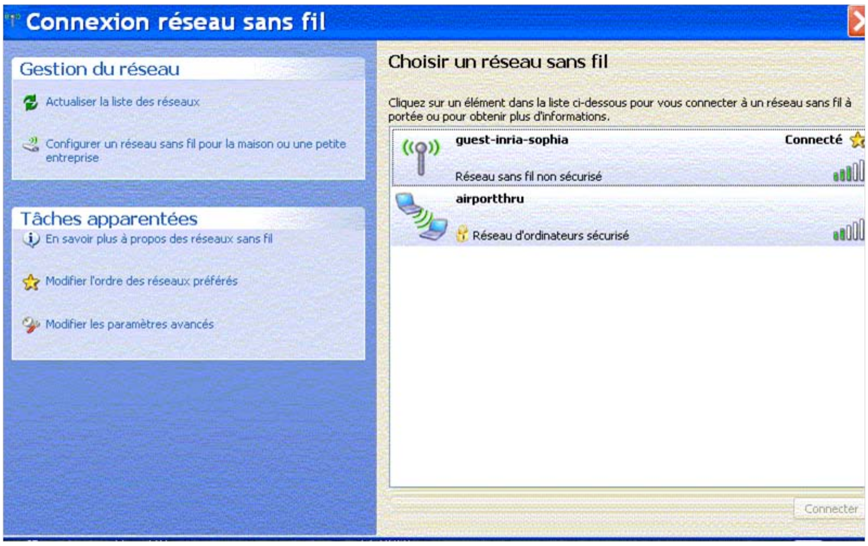


Figure 1.6: The access point association problem: Information available after taking the initial decision

period. it may discover the quality of service offered by an operated only after taking the decision of whicih of the service providers to connect to.

In this game the decisions may depend on the pricing strategy of each service provider as well as on the quality of its service. The latter may be unknown, and become available only after taking where as the former may become available

1.6 Exercise for the end of the course

Formulate and solve the association problem of Section 1.5.

Figure 1.7: The association problem of choice of technology: Information available when taking a decision

Chapter 2

Motivating Example: Transport Protocols

Non-cooperative Game theory tries to predict stable outcomes of competition. There may be various ways to understand a prediction of the outcome of competition, and the standard concept of Nash equilibrium is one of them. Before presenting the definition of the Nash equilibrium we mention other alternative concepts that or tests that have been used to predict future evolution of competition. We then intnroduce concepts from game Theory that predict the evolution of competition and provide the motivation for using them

2.1 Competition between protocols, the indifference property

Transport protocols are mechanims that are used for transmitting data in the Internet. We focus on TCP (Transmission Control Protocol) which is a family of protocols that carry a vast majority of the packets in the Internet. Its objectives are

- (a) Detecting packets that are lost in the network, i.e. that did not reach the destination within a given time limit. Loss of packets may occur during congestion at buffers in routers along the path of the packets from the source to the destination. A congestion is a period during which the rate of arrival of packets to a buffer exceeds the rate at which the buffer can handle them. During such periods the number of buffered packets grows until the buffer fills and overflows, which causes losses. We note that losses may also be caused by noisy links and are common in wireless communication.

- (b) Retransmitting those packets from the source to the destination.

- (c) Adapting the transmission rate to the available bandwidth so as to avoid congestion. All variants of TCP are based on increasing gradually the transmission rate until congestion indications are received (in general these indications are simply the detection of a packet loss). This then triggers a decrease in the transmission rate.

Consider two types of protocols, A and B. We would like to project from measurements how will the future Internet look like, assuming that consumers will prefer to use protocols that perform better. Will it be composed only of aggressive or of friendly protocols? Or a mixture of them? In the latter case, what will be the fraction of aggressive protocols?

2.2 Predicting the protocoles that will dominatne Future Internet

There have been various approaches to answer the above questions.

A1(u) The isolation test See how well the protocol performs if everyone uses the friendly protocol only. Then imagine the world with the aggressive TCP only. Compare the two worlds. The version u for which users are happier is the candidate for the future Internet. Example for this prediction approach is [7][1]

A2(u) The Confrontation test Consider interactions between agressive and pieceful sessions that share a common congested link. The future Intrnet is predicted to belong to the transport protocol of version u if u performs better in the interaction with v.

A3. Combined approach: Assume that a version u does better than v under both the isolation test as well as the confrontation test. We then call this version a strong TCP version and predicn that it will dominate future Internet. If there is no strong of TCP, we predict that both versions will co-exist. A weak TCP version is one that fails in both tests.

A4. Comparative test: Assume that everyone uses a version u of TCP and that one session starts using a version v instead of u. The comparative approach would choose u as a candidate for dominating the future internet if the performance of the deviating TCP (that uses version v) is strictly inferior to that of u. We then call u a strict Nash equilibrium in pure strategies. It is called a Nash Equilibrium in pure strategies if the inequality is non-strict.

A5. Indifference approach: Assume that some fraction p of the population uses version $T1$ and a fraction $1-p$ uses version $T2$. Assume that p is such that **the average performance of a protocol is the same under both u and v.** This is called the indifference test. When it holds then we say that the g ame has a Nash equilibrium $(p, 1-p)$ in mixed strategies.

2.3 The TCP Game: General Examples

Let us introduce a two player game with Sally, the first player, and Van, the second one. We assume that a bottleneck is sharead by two TCP coonnections. Each player controls one of these connections, and has to choose the version of the protocole among the two versions T1 and T2. We write the utilities in a matrix form where Sally chooses a row (indexed by T1 and T2) and Van chooses a column (indexed by T1 or T2).

Sally \ Van	$T1$	$T2$
$T1$	$2,2$	$10,0$
$T2$	$0,10$	$3,3$

Figure 2.1: Game in Example 1

Sally \ Van	$T1$	$T2$
$T1$	$3,3$	$0,1$
$T2$	$1,0$	$2,2$

Figure 2.2: Game in Example 2

Sally \ Van	$T1$	$T2$
$T1$	a,a	b,c
$T2$	c,b	d,d

Figure 2.3: General symmetric two by two matrix game

[1]The author of this references writes in the concluding Section "The last issue, which was not addressed in this paper, concerns the deploying of TCP Vegas in the Internet. It may be argued that due to its conservative strategy, a TCP Vegas user will be severely disadvantaged compared to TCP Reno users, ... it is likely that TCP Vegas, which improves both the individual utility of the users and the global utility of the network, will gradually replace TCP Reno."

Example 1. *Consider the payoffs as given in Figure 2.1. Version $T1$ is seen to dominate version $T2$ in the Isolation Test, providing a utility of 3 units where as version $T2$ gets only 2 units in isolation. In contrast, the confrontation test as well as the comparative test select version $T1$.*

Example 2. *Consider next the situation in Figure 2.2. T1 is a strong strategy - it satisfies both the Isolation and Confrontation Tests, where as T2 is a weak strategy - it does not satisfy any of these. Yet both satisfy the comparative test and are thus Nash equilibria in pure strategies.*

Remark 2. *A word on the notation. The examples that we encounter in this Section are two by two symmetric matrix game. In the matrix representations that we had in the previous examples, the first chooses a row and the second chooses the column. The first of the two numbers in the correspondingn entry of the matrix is then the utility for player 1 and the second - that of player 2. Of course, row i (respectively, column j) corresponds to choosing T_i (respectively, T_j) by player 1 (resp. player 2).*

In two by two matrix games we shalll use sometime for simplicity T (Top) and B (bottom) to describe choices of rows, and L (left) and R (right) for the choice of column.

We restrict here to symmetric games. It is thus sufficient to have only one number in each of entries of the matrix G that represents the game. In that case, we define G_{ij}, the ij-th entry of the matrix G, to be the utility of player 1 when it uses action i and the other player uses action j. Due to symmetry, the utility of player 2 is then given by G_{ji}.

Remark 3. *The generic symmetric two by two matrix game form is given in Figure 2.3. If all parameters (a, b, c) are different, then one classifies the generic game into three classes.*

- *The case $(a - c)(d - b) < 0$ is called the Prisoner's Dilema game. It has one single pure equilibrium which is said to be dominating. Example 2.1 falls within this category.*
- *The case $a > c, d > b$ is called a coordination game. It has two pure equilibria: a, a and d, d, and one mixed equilibrium. Example 2.2 falls into this category.*
- *The case $a < c, d < b$. The game has one single equilibrium which is mixed. This game can be seen to satisfy the indifference criterion. The game is known as the Hawk and Dove game.*

2.4 The TCP Game: New Reno Vs Scalable TCP

Consider two types of protocols:

- (i) *aggressive*, which try to rapidly grab as much bandwidth as possible, and
- (ii) *friendly*, which are much slower to grab extra bandwidth.

More specifically, we shall consider the NR (New Reno) and Sc (Scalable) versions of TCP. Sally and Van have to choose which of the two TCP protocols they will use. They first simulate in ns-2 [8] (and repeat their experiment several times) and discover that when a bottleneck is shared by two TCP connections then

- NR vs NR: share fairly the link capacity. If the buffers are well dimensioned then each one's throughput is close to half the link's capacity.
- Sc vs Sc: Again, if the buffers are well dimensioned then the sum of throughputs of the connections will be the available bandwidth (speed at which packets leave the buffer). By symmetry, each one will receive half the bandwidth. However, unless the connections start at the same time, it will take a very long time till they share the buffer fairly, unlike NR vs NR where fairness is achieved very fast. We conclude that in Sc vs Sc there is a short term unfairness. We assume that users are unhappy when treated unfairly, and represent this with some cost δ.

- Sc vs NR: The share of NR denoted by α is smaller than that of Sc denoted by $1-\alpha$. Thuss $\alpha < 0.5$. There are no fairness issues here other than the fact that the shares are shared unfairly.

We summarize this in Fig. 2.4. We shall assume below that $a + d > 0.5$.

		Van	
		H	D
Sally	H	$0.5-\delta$	$1-\alpha$
	D	α	0.5

Figure 2.4: Aggressive (H) versus friendly (D) TCPs

According to Remark 3, this is the "Hawk-Dove" game. It has a unique mixed equilibrium between aggressive (Hawk) behavior and the peaceful (Dove) one. We shall use H for Sc TCP and D for NR.

We observe that NR wins in the isolation test and that Sc wins in the confrontation test. Neither versions wins the comparison test, but the indifference principle holds. There exists a unique equilibrium which is mixed.

Reccall that we had assumed that $0.5 < a + d$. If we assume the opposite inequality to hold then the game becomes a prisoner's dilema game with a single pure equilibrium.

2.5 Predicting the Evolution of protocols

After introducing different tests and criteria and after introducing the TCP game, we are ready to examine the question of - which of the criteria are relevant in predicting the structure of the future Internet.

We adopt in this book game theoretic answers to the above question.

Game theory is concerned with predicting both the outcome (equilibrium) as well as the dynamics of competition. The Nash equilibria are the candidates for the possible stationary points of competition dynamics. This will be discussed in more details at later chapters. Nash equilibria have the property that if both players start initially at equilibrium and then react to each other by using say, at a round robin way, their optimal response to the other player, then the players remain at that equilibrium. In that sense the Nash equilibrium predicts potential stable outcomes of competition. Here we mean by Stable that no player can strictly do better by deviating unilaterally and playing another action than the equilibrium one.

Note: Nash equilibrium does not state what can or cannot happen when more than one decision maker changes their strategy (route) simultaneously.

Much later after the introduction of Nash equilibria, evolutionary equilibria notions appeared as well. The question of convergence when starting away from equilibria as well as convergence in the presence of mutations have been studied [?].

[?] showed under some technical conditions, that even if mutations continue to appear and we start away from equilibrium, we shall converge to a dominant equilibrium, if sucuh exists. Here, an equilibrium is said to be dominant if it outperforms any other equilibrium.

We saw that competition between two TCP versions can be formulated as either one of the following games: the prisoner's dilema, the Hawk-Dove game or the coordinated game. The two first ones have a unique equilibrium. Thus from the evolutionary game literature (e.g. [?]), we may predict that in the Hawk-Dove game (that arises in the competition between NR and Sc versions of TCP) both versions will coexist. We shall compute later in the chapter the proportions that each policy is used. If the game were to follow the prisoner's dilema then example then the future Internet would consist of one dominating version of TCP. In the coordination case we would expect to have convergence to the dominating equilibrium.

2.6 Background on the evolution of transport protocols

Today, NR turns out to be a very popular version of TCP. Although Sc is more aggressive than NR, it is more friendly than the TCP used more than twenty years ago, which did not at all adapt its rate to the congestion. At that time the Internet suffered from "congestion collapse" periods during which the throughput was very small and during which many packets were lost and had to be retransmitted. Van Jacobson then invented the first TCP that had an adaptive reaction to congestion, called TAHOE (TA). With TA, the congstion collapse was prevented. Thus in a game between TAHOE and the previous version, TA wins the isolation test and looses the confrontation test.

Reno version came after Tahoe. Reno is slightly more aggressive than Tahoe; in a game against Tahoe it wins both the isolation as well as the confrontation tests. It thus dominated Tahoe, which is thus not in use any more. NR is an improved version of Reno and in a game against Reno, NR dominates.

Much has been written on the comparison between NR and the Vegas version of TCP. The latter is more friendly than NR; it detects congestion not only through losses but also by measuring the end-to-end delay. In a game against NR, Vegas wins the isolation test but looses in the confrontation one.

In what follows, we shall provide more precise definitions of the equilibria and of the strategies.

Chapter 3

Equilibria

3.1 Mixed strategies

Assume that neither actions is dominant in the TPC game. The game approach predicts that both versions will coexist. The fraction of each is computed so that the average performance of a protocol is the same under both actions.

Consider again the game depicted in Fig. 2.4. Let this fraction of actions be p and $1-p$. Take $a=0$. If Sally plays NR then her utility is

$$U(NR,p) = \underbrace{(0.5-d)} \times p + 1 \times (1-p) = 1-(0.5+d)p.$$

Probability that Van plays NR	Payoff for Sally $M_{NR,NR}$	Probability that Van plays Sc	Payoff for Sally $M_{NR,Sc}$

Else her utility is

$$U(Sc,p) = 0.5(1-p).$$

Equating $U(Sc,p) = U(NR,p)$ we get

$$p = \frac{1}{1+2d}.$$

This is the fraction of sessions that will use Sc. It goes to 1 as $d \to 0$.

3.2 Nash Equilibrium

We introduced game theory in a context of competition between several actions that are used by players that interact with each other. The game allows to predict wheather in the "future Internet" an action would dominate the other or wheather several actions would coexist, and if so, with what probabilities.

We did not justify this definition; in fact there is no reason to assume that a competition between two versions of protocoles would lead the fraction of each one of them to converge to some limit value. If we have convergence with the properties that we described, then we say that the game has a stable equilibrium. In the case of dominating strategies, the equilibrium was obtained by one version of TCP, so we called it a "pure" equilibrium. Otherwise it was described as a fraction that uses each verisson. We called this an equilibrium in "mixed" strategies.

We note the following cruicial property of an equilibrium. At equilibrium, the stgrategy of each player turned out to coincide with the best performance it could achieve when the other players use the equilibrium strategy. Indeed, in case of a dominating action, this was true by definition (of dominating action). In the case of mixed action, this holds by the simple fact that at equilibrium, any strategy is optimial if the other chooses the equilibrium strategy. (This was the indifference rule).

Let $\mathcal{A}^i$ be the finite set of actions (or of strategies) available to player i. We call $\mathbf{a} = (a_1, ..., a_I)$ a multi strategy where a_i inf $\mathcal{A}^i$ and denote by $\mathcal{A}$ the set of all multi-strategies. For a multi-strategy $a \in \mathcal{A}$ and an action $a' \in \mathcal{A}^i$, define $(a'|a_{-i})$ to be the multistrategy obtained from a by a deviation of player i from a^i to a'.

We are now in measure of defining matrix games.

3.2.1 Nash Equilibrium in Pure Strategies

We start with, as input, the tensor (a tensor is the extension of a matric to dimension larger than 2) that gives the utility for a player (i.e. the way the player evaluats the performance) of all possible combinations of actions of all players. Let there be N_i actions available to player i, $i = 1, 2, ..., I$. Let $U^i(k, l)$ be the utility or payoff for player i when player i chooses action k and the other player chooses action or strategy l.

In our example we had a symmetric game: $U^1(j, k) = U^2(k, j)$. We thus spcified only one utility function. More generally there will be as many utility matrices as there are players.

A matrix game is defined as a non-cooperative optimization performed by a number of players, each with a utility function as described. The basic solution concept of this optimization is the equilibrium.

Definition 1. *We say that a multi-strategy $a^* \in \mathcal{A}$ is a "pure" eqeuilibrium if for every player i and any $a \in \mathcal{A}^i$,*

$$U^i(a^*) \geq U^i(a_i, a^*_{(-i)})$$

We note that the TCP game does not have a pure equilibrium.

3.2.2 Nash Equilibrium in Mixed Strategies

Let $\Delta(\mathcal{A}^i)$ be the set of probability measures over the set of actions of player i. These are the mixed strategies of player i. We define a mixed multi-strategy as a vector of dimension I containing one mixed strategy per player. We define the utility corresponding to a mixed strategy pair (p, q) as

$$U(p, q) = \sum_{j=1,...N^1} \sum_{k=1,...N^2} p(j)U(j, k)q(j)$$

which we write in vector n otation as $p^T U q$.

Definition 2. *We say that a multi-strategy $p^* \in \Delta\mathcal{A}$ is a "mixed" eqeuilibrium if for every player i and any $p \in \mathcal{A}^i$,*

$$U^i(p^*) \geq U^i(p_i, p^*_{(-i)}).$$

We shall show later that

Theorem 1. *A matrix game always has a mixed equilibrium.*

We note that we identify a pure strategy as a (degenerate) special case of a mixed strategies.

There are various ways to obtain equilibria of matrix games, see [9, 10][1]

3.3 2 player zero-sum games: Upper and Lower Value

This is a special case of $U(a, b) := U^1(a, b) = -U^2(a, b)$.

Zero-sum games model conflicting situations. It is also used for robust control: a design against a worst possible situation is done by imagining another player with opposite objectives.

Note: any equilibrium is Pareto optimal: we cannot increase the utility of one without decreasing that of another one.

[1] See http://www.maths.lse.ac.uk/Personal/stengel/TEXTE/agt-stengel.pdf as well as http://www.maths.lse.ac.uk/Personal/stengel/TEXTE/nashsurvey.pdf

3.3.1 Definitions

Define the Lower Value:

$$\underline{J} := \max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}} U(a, b)$$

Let a^* be the argument achieving the maximum, i.e.

$$\min_{b \in \mathcal{B}} U(a^*, b) = \overline{J}$$

We call it the maxmin solution for player 1, or the maxmin optimal strategy.

Define the Upper Value:

$$\overline{J} := \min_{b \in \mathcal{B}} \max_{a \in \mathcal{A}} U(a, b)$$

Let b^* be the argument achieving the minimum, i.e.

$$\max_{a \in \mathcal{A}} U(a, b^*) = \overline{J}$$

We call it the minmax solution, or the maxmin optimal strategy.

We have the following relation:

$$\underline{J} = \min_{b \in \mathcal{B}} U(a^*, b) \leq U(a^*, b^*) \leq \max_{a \in \mathcal{A}} U(a, b^*) = \overline{J}$$

If

$$\max_{a \in \mathcal{A}} U(a, b^*) = \min_{b \in \mathcal{B}} U(a^*, b)$$

then (a^*, b^*) is called a saddle point and $J = \overline{J} = \underline{J}$ is called the value of the game.

A zero-sum matrix game does not necessarily have a value over the pure strategies.

We next extend the above definition to the mixed strategies. Recall the definition $U(p, q) = p^T U q$, and consider the matrix game over the mixed strategies. We define the the maxmin, minmax and saddle point strategies for the set of mixed strategies in the same way as we did in for the pure strategies.

3.3.2 Saddle-point theorems

Theorem 2. *A zero-sum matrix game over mixed strategies always has a saddle point.*

Usually the maxmin and minmax value of a game are defined as those corresponding to pure strategies, where as the value of the game is defined as corresponding to mixed strategies!

Remark 4. *One can interpret* $\max_a \min_b U(a, b)$ *as if there is an order between the moves: first player 1 chooses an action a and then player 2, knowing the choice of player a, minimizes* $U(a, b)$*. Indeed, note that when writing* $\min_b U(a, b)$*, player 2 makes a choice that depends on a. Denote by* $b(a)$ *that choice. Player 1 however maximizes the function* $\min_b U(a, b)$ *or eqeuivalently it maximizies* $U(a, b(a))$*.*
When playing over mixed strategies then $\max_p \min_q U(p, q)$ *has the interpretation that theh choice of player 2 may depend on the choice of probability distribution of player 1, but not on its realization!*

In unconstrained matrix games, as well as in matrix games with orthogonal constraints (see [11]), a saddle point in mixed strategies always exists, and the saddle point payoff is the value of the game. Moreover, if (x_1, y_1) and (x_2, y_2) are two saddle points, then (x_1, y_2) and (x_2, y_1) are also saddle points. More generally, the following holds for zero-sum games (e.g., [12, p. 126]):

Lemma 1. (*Minmax Theorem*)
Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be convex subsets of linear topological spaces, where $\mathcal{S}_2$ is compact. Consider a function $U : \mathcal{S}_1 \times \mathcal{S}_2 \to \mathbb{R}$ such that
– for each $x \in \mathcal{S}_1$, $y \to U(x, y)$ is convex and lower semi-continuous; and
– for each $y \in \mathcal{S}_2$, $x \to U(x, y)$ is concave.
Then there exists some $y_* \in \mathcal{S}_2$ such that

$$\inf_{y\in\mathcal{S}_2}\sup_{x\in\mathcal{S}_1} U(x,y) = \sup_{x\in\mathcal{S}_1} U(x,y_*) = \sup_{x\in\mathcal{S}_1}\inf_{y\in\mathcal{S}_2} U(x,y).$$

We conclude that under the conditions of the lemma, if $\mathcal{S}_1$ is compact as well then a saddle point exists. This Lemma also holds for games with orthogonal constraints.

3.3.3 Linear Programming Solution

The value of a matrix game satisfies:

$$J = \min_q J(q)$$

where

$$J(q) = \max_p p^T U q.$$

Then

$$J(q) \geq \sum_b U(a,b)q(b) \quad \forall\ a \tag{3.1}$$

If any other constant $J'(q)$ satisfied the inequality then it would also satisfy

$$J'(q) \geq \max_a U(a,q)$$

or equivalently, $J' \geq J$. Hence $J'(q)$ is the smallest constant that satisfies (3.1) (since it achieves it with equality. We conclude that J is the smallest solution of $J'(q) \geq U(a,q)$. Hence J is the solution of

Minimize J over all J and q under the constraints $J \geq U(a,q)\ \ \forall a$, where $\forall a, q(a) \geq 0, \sum_a q(a) = 1$.

To solve a zero-sum game in maple, use the code:

```
with(Optimization):
LPSolve(J,{J >=q1* U11 + q2*U12 , J >=q1* U21 + q2*U22 , q1+q2=1 , q1 >= 0 ,
q2 >= 0});
```

The second line can be replaced by

```
LPSolve(J,{J >=q1* U11 + q2*U12 , J >=q1* U21 + q2*U22 , q1+q2=1},
assume=nonnegative);
```

LPSolve understands this by default as a minimization problem. To overide this, add "maximize" before the last parenthesis.

	action 1.a	action 1.b
action 2.i	$2, 1$	$0, 0$
action 2.ii	$0, 0$	$1, 2$

Table 3.1: A matrix game [Aumann]

3.4 Corelated Equilibrium

3.4.1 Definition and properties

Consider the coordination game in Table 12.1. A player maximizes its **expected** utility. Player 1 chooses columns, Player 2 chooses rows.

There are two pure Nash equilibria: (1.a , 2.i) with the values (2,1), and (1.b , 2.ii) with values (1,2).

The randomized strategy (1/3,2/3) for player 1 and (2/3,1/3) for player 2 is (the unique) symmetric equilibrium. The corresponding values are

$$(2/9) \times 1 + (2/9) \times 2 = 2/3 \quad \text{for each player}$$

Suppose that an arbitrator **suggests** suggests to both either ($1.a$, $2.i$) (w.p. 1/2) or ($1.b$, $2.ii$) (w.p. 1/2).

If the players follow the advise they obtain in expectation 3/2 each. Cannot be obtained without coordination.

The game is still non-cooperative: no "binding contract".

Here, the correlated equilibrium does not dominate the pure Nash equilibria.

There are cases where it dominates all other Nash equilibria.

The coordinator does not have to know anything about the game: all it has to do is flip a coin and send signals.

Correlation is needed for **coordination**. Useful not just in games but also in team problems (a common objective).

Ref: R. J. Aumann, ***Journal of Mathematical Economics*****, 1974.**

3.4.2 Applications: ALOHA

Players: N mobile stations transmitting to a Base Station.

If more than one mobile attempts transmission at the same time, the packets are lost.

Strategies: transmission probabilities.

Objective: maximize throughput, constraints on average power.

Correlation mechanism: Each mobile has a serial number within $1, ..., K$ where $K << N$. BS transmits a random number k in $1, ..., K$.

Candidate equilibrium: Mobiles with $k' \neq k$ will not attempt transmission. The others will transmit with some probability (to determine).

Ref (without correlation): E. A., R. El-Azouzi and T. Jimenez, Computer Networks, 2004.

3.5 Concave games and Nash equilibrium

Chapter 4

Basic Examples

4.1 Multiple-access game over a collision channel, Transmit or Wait

Consider two players having a packet to transmit over a collision channel. Each player may transmit or wait. If both transmit simultaneously then the packets are both lost. The payoff for each player is one if the transmission is successful. We obtain the matrix game in Figure 4.6.

Transmitting is a dominant strategy for both player. It leads to zero throughput and zero utility.

(a) Original Version: Player 1 (rows), Player 2 (columns)

	H	D
H	0	1
D	0	0

(b) Version with energy cost: Player 1 (rows), Player 2 (columns)

	H	D
H	$-E$	$1-E$
D	0	0

Figure 4.1: The Multiple-Access Game

We add next an energy cost E if a player transmits. We obtain the matrix game in Figure 4.1(b).

If $E \geq 1$ then $r = 0$ is the unique equilibrium.

Assume $0 < E < 1$. We see that the game is of the third type in Remark 3. It is thus a H-D game with a single mixed equilibrium given by

$$r = 1 - E$$

This also leads to zero utility but now the equilibrium throughput is

$$Thp = r(1-r) = E(1-E)$$

This is maximizied at $E = 1/2$ which gives an equilibrium throughput of $1/4$.

We note that a correlated equilibrium would lead a better throughput. The arbitrator sends to the mobiles the outcome of a randomization that decides who of them will transmit. A total throughput of 1 can be achieved.

Why did the utility at equilibrium remain zero? Due to the indifference property: at equilibrium, player 1 is indifferent between 1st and 2nd row. The utility for the 2nd row is always 0 independently of other players.

		Player 2 H	Player 2 D
Player 1	H	Q_{11}	Q_{12}
	D	Q_{21}	Q_{22}

Figure 4.2: High power (H) versus low power (D)

Player i can transmit with power $p_i \in \mathcal{A} := \{P_H, P_L\}$. We now consider capture. Let Q_{ij} be the probability that the transmission of player 1 is successful when using power p_i and the other one uses p_j. Example: $Q_{HH} = p$, $Q_{DD} = q < p$, $Q_{HD} = 1$, $Q_{DH} = 0$. Dominating strategy: H. Gives a throughput of p to each mobile.

4.2 At what Power to Transmit? The Capture Phenomenon

How to compute capture probability: we have the following expressions for the bit error probability as a function of the modulation [13] (numerical examples based on these formulas can be found in [14, 15, 13]):

$$p_e(SINR) = \begin{cases} \frac{1}{2} erfc(\sqrt{\kappa \cdot SINR}) & \text{for GMSK} \\ \frac{1}{2}\exp\left(-SINR\right) & \text{for DBPSK} \\ \frac{1}{2}\exp(-\frac{1}{2} \cdot SINR) & \text{for GFSK} \\ \frac{1}{2} erfc(\sqrt{SINR}) & \text{for QPSK} \\ \frac{3}{8} erfc(\sqrt{\frac{2}{5} \cdot SINR}) & \text{for 16-QAM} \\ \frac{7}{32} erfc(\sqrt{\frac{4}{21} \cdot SINR}) & \text{for 64-QAM} \end{cases}$$

where κ is a constant (that depends on the amount of redundancy in the coding and on the frequency band), and where $erfc$ is the complementary error function given by

$$erfc(x) = \frac{2}{\pi} \int_x^{\infty} e^{-\zeta^2} d\zeta.$$

In the absence of redundancy this gives the following expression for f of a packet of N bits provided that the bit loss process is independent

$$f(SINR) = (1 - p_e(SINR))^N$$

		Player 2 H	Player 2 D
Player 1	H	$p - E$	$1 - E$
	D	0	q

Figure 4.3: High power (H) versus low power (D)

We add a cost E for using the high power. H remains a dominant strategy as long as $E < 1 - q$. For such E, the uhtility is $p - E$.

Note: The utility at equilibrium can decrease or increase when decrease the entries of the matrix. This is not the case in optimization, nor in zero-sum games.

4.3 Coordination games over a collision channel

Simple Motivating Example Consider the following basic example. There are two mobiles $i = 1, 2$ and two independent channels $j = 1, 2$. Each mobile transmits at the same time one packet: Mobile i transmits a packet over channel i with probability p_i and with probability $1-p_i$ over the other channel. A packet is successfully transmitted if it is the only one that uses the channel. Thus the transmission success probability of mobile i is

$$U_i(p) = p_i p_j + (1-p_i)(1-p_j), \qquad j \neq i$$

Mobile i wishes to maximize the probability U_i of successful transmission of its packet.

The policy that assigns a dedicated channel to each mobile (i.e. $p_1 = p_2 = 1$ or $p_1 = p_2 = 0$) is obviously optimal: it involves no collisions and the success probability is one. It is also a Nash equilibrium. However it requires coordination or synchronization in order to assign each channel to a different mobile.

The symmetric policy $p_1 = p_2 = 0.5$ turns out to be an equilibrium; if mobile i uses $p_i = 0.5$ then no matter what p_j mobile j $(i \neq j)$ chooses, it will have the same success probability of $1/2$. Thus no mobile can benefit by unilaterally deviating from $p = 1/2$, so it is an equilibrium.

Note that if mobile i had only one option, that of choosing channel i, then the innefficient equilibrium would not occur. This is a feature similar to the inefficiency we have in the prisoner's dilemma or in the Braess' paradox in which eliminating some options for the players can result in better performance to every one. Yet if we wanted to implement this idea in our context and create mobiles with only one channel, then we would face again a synchronization problem. If half of the mobiles have built in technology for accessing one channel and the other half can only access the other channel, then two randomly selected mobiles will still be using the same channel with probability half.

The model and main result Consider 2 mobiles and 2 base stations. The base stations use, each one, an independent channel (for example, each one uses another frequency). We shall assume that mobile i has a good radio channel with base station i and a bad one with station $j \neq i$. More precisely, let h_{ij} be the gain between mobile i and base station j.

Let $SINR_i$ denote the Signal to Interference and Noise Ratio corresponding to the signal received from mobile i at the base station to which it transmits. Each mobile has two pure strategies: γ, β where γ means transmitting on its good channel and β on its bad one. Then

$$SINR_1(u) = \begin{cases} \dfrac{h_{11}P_1}{N_o} & \text{if } u = (\gamma,\gamma) \\[2ex] \dfrac{h_{12}P_1}{N_o} & \text{if } u = (\beta,\beta) \\[2ex] \dfrac{h_{11}P_1}{N_o + h_{21}P_2} & \text{if } u = (\gamma,\beta) \\[2ex] \dfrac{h_{12}P_1}{N_o + h_{22}P_2} & \text{if } u = (\beta,\gamma) \end{cases}$$

$$SINR_2(u) = \begin{cases} \dfrac{h_{22}P_2}{N_o} & \text{if } u = (\gamma,\gamma) \\[2ex] \dfrac{h_{21}P_2}{N_o} & \text{if } u = (\beta,\beta) \\[2ex] \dfrac{h_{22}P_2}{N_o + h_{12}P_1} & \text{if } u = (\beta,\gamma) \\[2ex] \dfrac{h_{21}P_2}{N_o + h_{11}P_1} & \text{if } u = (\gamma,\beta) \end{cases}$$

		Player 2	
		β	γ
Player 1	β	$\left(\frac{h_{12}P_1}{N_o}, \frac{h_{21}P_2}{N_o}\right)$	$\left(\frac{h_{12}P_1}{N_o + h_{22}P_2}, \frac{h_{22}P_2}{(N_o + h_{12}P_1)}\right)$
	γ	$\left(\frac{h_{11}P_1}{N_o + h_{21}P_2}, \frac{h_{21}P_2}{N_o + h_{11}P_1}\right)$	$\left(\frac{h_{11}P_1}{N_o}, \frac{h_{22}P_2}{N_o}\right)$

Figure 4.4: SINR values of the two mobiles

Here N_o is the thermal noise at each base station and P_i is the fixed transmission power of mobile i.

Under many modulation schemes the probability of a successful transmission of a packet is known to be a monotone increasing function of the $SINR$ [16]. We thus assume that mobile i has a success probability given by $f_i(SINR_i)$. Define

$$A := f_1\left(\frac{h_{11}P_1}{N_o},\right) \quad B := f_1\left(\frac{h_{11}P_1}{N_o + h_{21}P_2}\right)$$

$$C := f_1\left(\frac{h_{12}P_1}{N_o + h_{22}P_2}\right) \quad D := f_1\left(\frac{h_{12}P_1}{N_o}\right).$$

$$a := f_2\left(\frac{h_{22}P_2}{N_o},\right) \quad b := f_2\left(\frac{h_{22}P_2}{N_o + h_{12}P_1}\right)$$

$$c := f_2\left(\frac{h_{21}P_2}{N_o + h_{11}P_1}\right) \quad d := f_2\left(\frac{h_{21}P_2}{N_o}\right).$$

The mobiles are thus faced with the following matrix game:

	action γ	action β
action γ	A, a	B, c
action β	C, b	D, d

Theorem 3. *There are exactly three equilibria; the two pure equilibria: (γ, γ) and (β, β), and a mixed one in which player 1 and 2 select γ with probabilities:*

$$X^* = \frac{D - B}{A + D - B - C}, \quad Y^* = \frac{d - b}{a + d - b - c}.$$

Proof. We note that $B < D$, $b < d$, $C < A$ and $c < a$. The game is thus a standard coordination game (see *http://en.wikipedia.org/wiki/Coordination_game*) [17] for which the result is well known. ■

The mixed equilibrium is characterized by the indifference property: when a mobile uses its mixed equilibrium policy then the other player is indifferent between γ and β.

The utility of mobile 1 and 2 at the mixed equilibrium are given by

$$U_1^* = AY^* + B(1 - Y^*), \quad U_2^* = aX^* + b(1 - X^*).$$

Consider now the symmetric case $(A = a, B = b, C = c, D = d)$. Then we get at the mixed equilibrium:

$$U^* = \frac{(a - c)(d - c) + c(a + d - b - c)}{a + d - b - c} = \frac{ad - cb}{a + d - b - c}$$

4.4 The congestion Game

4.5 The Hawk and Dove (HD) Game

Consider a large population of animals. Occasionally two animals find themselves in competition on the same piece of food. An animal can adopt an aggressive behavior (Hawk) or a peaceful one (Dove). The matrix in Fig. 13.2.1 presents the fitness of player I (some arbitrary player) associated with the possible outcomes of the game as a function of the actions taken by each one of the two players. We assume a symmetric game so the utilities of any animal (in particular of player 2) as function of its actions and those of a potential adversary (in particular of player 1), are the same as those player 1 depicted in Figure 13.2.1. The utilities (i.e. fitness) represent the following:

An encounter D–D results in a peaceful, equal-sharing of the food which translates to a fitness of 0.5 to each player.

An encounter H–H results in a fight in which with equal chances, one or the other player obtains the food but also in which there is a positive probability for each one of the animals to be wounded. Then the fitness of each player is 0.5-d, where the 0.5 term is as in the D–D encounter and the $-d$ term represents the expected loss of fitness due to being injured.

An encounter H–D or D–H results in zero fitness to the D and in one unit of utility for the H that gets all the food without fight.

		Player II	
		H	D
Player I	H	$0.5-d$	1
	D	0.5	0.5

Figure 4.5: A H–D game in matrix form

		Player II	
		H	D
Player I	H	$A11$	$A12$
	D	$A21$	$A22$

Figure 4.6: Generalized H-D game

PPP

One can think of other scenarios that are not covered in the original H–D game, such as the possibility of a Hawk to find the Dove, in a H–D encounter, more delicious than the food they compete over. A generalized version [18] of the HD game given in Figure 4.6 is characterized by $A11 < A22 < A12$ and $A21 < A22$. In that case,

1. if $A11 > A21$ then the pure strategy H is the unique ESS,
2. If $A11 < A21$ then there is a unique ESS $p = (p_L, p_H)$, it is a mixed strategy given by $p_H = u/(u+v)$ where $Aij = J(i,j),\ i,j \in \{H,D\},\ u = A12 - A22,\ v = A21 - A11$.

Remark 5. (i) Note that there are no settings of parameters for which the pure strategy D is an ESS in the H–D game (or in its generalized version).
(ii) In case 2 above, the strategies (H,D) and (D,H) are pure Nash equilibria in the matrix game. Being asymmetric, they are not candidates for being an ESS according to our definition. There are however contexts in which one obtains non-symmetric ESS, in which case they turn out to be ESS.

4.6 TCP over wireless

During the last few years, many researchers have been studying TCP performances in terms of energy consumption and average goodput within wireless networks [19, 20]. Via simulation, the authors show that the TCP New-Reno can be considered as well performing within wireless environment among all other TCP variants and allows for greater energy savings. Indeed, a less aggressive TCP, as TCP New-Reno, may generate lower packet loss than other aggressive TCP. Thus the advantage of an aggressive TCP in terms of throughput could be compensated with energy efficiency of a more gentle TCP version. (In Section **??** we shall illustrate another consideration that affects the competition between TCP versions.) The goal of this section is to illustrate this point, as well as its possible impact on the evolution of the share of TCP versions, through a simple model of an aggressive TCP.

The model. We consider two populations of connections, all of which use AIMD TCP. A connection of population i is characterized by a linear increase rate α_i and a multiplicative decrease factor β_i. Let $\zeta_i(t)$ be the transmission rate of connection i at time t. We consider the following simple model for competition.

(i) The RTT (round trip times) are the same for all connections.

(ii) There is light traffic in the system in the sense that a connection either has all the resources it needs or it shares the resources with one other connection. (If files are large then this is a light regime in terms of number of connections but not in terms of workload).

(iii) Losses occur whenever the sum of rates reaches the capacity C: $\zeta_1(t) + \zeta_2(t) = C$.

(iv) Losses are synchronized: when the combined rates attain C, both connections suffer from a loss. This synchronization has been observed in simulations for connections with RTTs close to each other [21]. The rate of connection i is reduced by the factor $\beta_i < 1$.

(v) As long as there are no losses, the rate of connection i increases linearly by a factor α_i.

We say that a TCP connection i is more aggressive than a connection j if $\alpha_i \geq \alpha_j$ and $\beta_i \geq \beta_j$. Let $\overline{\beta}_i := 1 - \beta_i$. Let y_n and z_n be the transmission rates of connection i and j, respectively, just before a loss occurs. We have $y_n + z_n = C$. Just after the loss, the rates are $\beta_1 y_n$ and $\beta_2 z_n$. The time it takes to reach again C is

$$T_n = \frac{C - \beta_1 y_n - \beta_2 z_n}{\alpha_1 + \alpha_2}$$

which yields the difference equation:

$$y_{n+1} = \beta_1 y_n + \alpha_1 T_n = q y_n + \frac{\alpha_1 C \overline{\beta}_2}{\alpha_1 + \alpha_2}$$

where $q = \frac{\alpha_1\beta_2 + \alpha_2\beta_1}{\alpha_1 + \alpha_2}$. The solution is given by

$$y_n = q^n y_0 + \left(\frac{\alpha_1 C \overline{\beta}_2}{\alpha_1 + \alpha_2}\right) \frac{1 - q^n}{1 - q}.$$

HD game: throughput-loss tradeoff

In wireline, the utility related to file transfers is usually taken to be the throughput, or a function of the throughput (e.g. the delay). It does not explicitly depend on the loss rate. This is not the case in wireless context. Indeed, since TCP retransmits lost packets, losses present energy inefficiency. Since energy is a costly resource in wireless, the loss rate is included explicitly in the utility of a user through the term representing energy cost. We thus consider fitness of the form $J_i = Thp_i - \lambda R$ for connection i; it is the difference between the throughput Thp_i and the loss rate R weighted by the so called tradeoff parameter, λ, that allows us to model the tradeoff between the valuation of losses and throughput in the fitness. We now proceed to show that our competition model between aggressive and non-aggressive TCP connections can be formulated as a HD game. We study how the fraction of aggressive TCP in the population at (the mixed) ESS depends on the tradeoff parameter λ.

Since $|q| < 1$, we get the following limit $\overline{y}$ of y_n when $n \to \infty$:

$$\overline{y} = \frac{\alpha_1 C \overline{\beta}_2}{\alpha_1 + \alpha_2} \cdot \frac{1}{1 - q} = \frac{\alpha_1 \overline{\beta}_2 C}{\alpha_1 \overline{\beta}_2 + \alpha_2 \overline{\beta}_1}.$$

It is easily seen that the share of the bandwidth (just before losses) of a user is increasing in its aggressiveness. Hence the average throughput of connection 1 is

$$Thp_1 = \frac{1 + \beta_1}{2} \times \frac{\alpha_1 \overline{\beta}_2}{\alpha_1 \overline{\beta}_2 + \alpha_2 \overline{\beta}_1} \times C.$$

The average loss rate of connection 1 is the same as that of connection 2 and is given by

$$R = \frac{1}{T} = \left(\frac{\alpha_1}{\overline{\beta}_1} + \frac{\alpha_2}{\overline{\beta}_2}\right)\frac{1}{C} \text{ where } T = \frac{\overline{\beta}_1\overline{\beta}_2 C}{\alpha_1\overline{\beta}_2 + \alpha_2\overline{\beta}_1}$$

with T being the limit as $n \to \infty$ of T_n.

Let H corresponds to (α_H, β_H) and D to (α_D, β_D) such that $\alpha_H \geq \alpha_D$ and $\beta_H \geq \beta_D$. Then, for $i = 1, 2$, $Thp_i(H, H) = Thp_i(D, D)$. Since the loss rate for any user is increasing in $\alpha_1, \alpha_2, \beta_1, \beta_2$ it then follows that $J(H, H) < J(D, D)$, and $J(D, H) < J(D, D)$. We conclude that the utility that describes a tradeoff between average throughput and the loss rate leads to the HD structure.

The mixed ESS is given by the following probability of using H:

$$x^*(\lambda) = \frac{\eta_1 - \eta_2\lambda}{\eta_3} \quad \text{where}$$

$$\eta_1 = \left(\overline{\mu}\frac{1+\beta_1}{2} - \frac{1+\beta_2}{4}\right)C, \; \eta_2 = \frac{1}{C}\left(\frac{\alpha_1}{\overline{\beta}_1} - \frac{\alpha_2}{\overline{\beta}_2}\right),$$

$$\eta_3 = C(\frac{1}{2} - \mu)\frac{\beta_1 - \beta_2}{2}, \; \mu = \frac{\alpha_2(\overline{\beta}_1)}{\alpha_2(\overline{\beta}_1) + \alpha_1(\overline{\beta}_2)}.$$

where $\overline{\mu} := 1 - \mu$. Note that η_2 and η_3 are positive. Hence, the equilibrium point x^* decrease linearly on λ. We conclude that applications that are more sensitive to losses would be less aggressive at ESS (Braess type paradoxes do not occur here).

For more details on this model, including the tradeoff between transient and steady-state behavior, we refer the reader to [22].

Chapter 5

Stackelberg Equilibrium and application to network neutrality

5.1 Definition and properties

Stackelberg game one introduces hierarchy between players. Actions are not taken anymore simultaneously. In a two level game, there are leaders that take actions first and then the others follow and take actions accordinigly. There is thus assymetry in the information: A follower decides knowing the leaders' current move.

This can be viewed as a game where the solution notion is a Nash equilibrium (so we "forget" the asymetry) but where the actions of the followers are in fact "strategies": they are functions of what the actions of the other player.

In a two player zero sum game, the two possible notions of Stackelberg give the upper and the lower value, respectively.

This remaining of the Chapter is based on [**?**].

5.2 Introduction to the network neutrality issue

Network neutrality is an approach to providing network access without unfair discrimination among applications, content, nor the specific source of traffic. What is discrimination and what is fair discrimination? If there are two applications or two services or two providers that require the same network resources and one is offered better quality of service (shorter delays, higher transmission capacity, etc.) then there is a discrimination. When is a discrimination "fair"[1]? A preferential treatment of traffic is considered fair as long as the preference is left for the user[2]. Internet service

[1] The recent decision on Comcast v. the FCC was expected by the general public to deal with the subject of "fair" traffic discrimination, as the FCC ordered Comcast to stop interfering with subscribers traffic generated by peer-to-peer networking applications. The Court of Appeals for the District of Columbia Circuit was asked to review this order by Comcast, arguing not only on the necessity of managing scarce network resources, but also on the non-existent jurisdiction of the FCC over network management practices. The Court decided that the FCC did not have express statutory authority over the subject, neither demonstrated that its action was "reasonably ancillary to the ... effective performance of its statutorily mandated responsibilities". The FCC was deemed, then, unable to sanction discrimatory practices on Internet's traffic carried out by american ISPs, and the underlying case on the "fairness" of their discriminatory practices was not even discussed.

[2] Nonetheless, users are just one of many actors in the net neutrality debate, which has been enliven throughout the world by several public consultations on new legislation on the subject. The first one, proposed in the USA (expired on 26/04/2010), was looking for the best means of preserving a free and open Internet. The second one, carried out in France (finishing 17/05/2010), asks for the different points of view over net neutrality. A third one is intended to be presented by the UE in the summer of 2010, looking for a balance on the parties concerned as users are entitled to an access the services they want, while ISPs and CPs should have the right incentives and opportunities to keep investing, competing and innovating. See [23, 24, 25].

providers (ISPs) may have interest in discrimination either for technological problems or for economic reasons. Traffic congestion has been a central argument for the need to discriminate traffic (for technological reasons) and moreover, for not practicing network neutrality, in particular to deal with high-volume peer-to-peer traffic. However, many ISPs have been blocked or throttled p2p traffic independently of congestion conditions.

There may be many hypothetical ways to violate the principle of network neutrality. Hahn and Wallsten wrote that network neutrality "usually means that broadband service providers charge consumers only once for Internet access, do not favor one content provider over another, and do not charge content providers for sending information over broadband lines to end users." (p. 1 of [26]) We therefore restrict our attention here to the practices of these types of network neutrality.

That net neutrality acts as a disincentive for capacity expansion of their networks, is an argument recently raised by ISPs. In [27] the validity of this claim was checked. Their main conclusion is that under net neutrality the ISPs invest to reach the social optimal level, while under-or-over investing is the result when net neutrality is dropped. In this case, ISPs stand as winners, while content providers (CP) move to a worst position. Users that rely on services that have paid the ISPs for preferential treatment will be better off, while the rest of the users will have a significantly worse service.

ISPs often justify charging content providers by their need to cover large and expensive amount of network resources. This is in particular relevant in the 3G wireless networks where huge investments were required for getting licenses for the use of radio frequencies. On the other hand, the content offered by a CP may be the most important source of the demand for Internet access; thus, the benefits of the access providers are due in part to the content of the CPs. It thus seems "fair" that benefits that ISP make of that demand would be shard by the CPs.

We find this notion of fair sharing of reveneus between economic actors in the heart of cooperative game theory. In particular, the Shapley value approach for splitting reveneus is based on several axioms and the latter fairness is on of them. Many references have advocated the use of the Shapley value approach for sharing the profits between the providers, see, *e.g.*, [28, 29]. We note however that the same reasoning used to support payments by access providers to content providers (in the context of can be used in the opposite direction. Indeed, many CPs receive third party income such as advertising revenue thanks to the user demand (eyeballs) that they create. Therefore, using a Shapley value approach would require content providers to help pay for the network access that is necessary to create this new income.

The goal is to study the impact of such side payments between providers on the utilities of all actors. More precisely, we study implications of one provider being at a dominating position so as to impose payments from the other one[3]. We examine these questions using simple game theoretic tools. We show how side payments may be harmful for all parties involved (users and providers).

Another way to favor a provider over another one is to enforce a leader-follower relation to determine pricing actions. We show how this too can be harmful for all.

5.3 Basic model: three collective actors and usage-based pricing

We consider the following simple model of three actors,

- the internauts (users) collectively,
- a network access provider for the internauts, collectively called ISP1, and
- a content provider and its ISP, collectively called CP2.

[3] In the European Union, such dominating positions in the telecommunications markets are controlled by the article 14, paragraph 3 of the Directive 2009/140/EC, considering the application of remedies to prevent the leverage of a large market power over a secondary market closely related.

In this section, the two providers are assumed peers; leader-follower dynamics are considered in Section 5.11 below. The internauts pay for service/content that requires both providers. Assume that they pay $p_i \geq 0$ to provider i (CP2 being $i = 2$ and ISP1 being $i = 1$) and that their demand is given by

$$D = D_0 - pd$$

where

$$p = p_1 + p_2 \geq 0, \;\; D \geq 0.$$

So, provider i's revenues are

$$U_i = Dp_i, \quad i = 1, 2.$$

5.4 Collaboration

The total price that the providers can obtain if they cooperate is maximized at $p_i = D_0/(4d)$. The total revenue per provider is then $U_i^{max} = D_0^2/(8d)$. The demand is then $D_0/2$.

5.5 Fair competition

If the providers do not cooperate then the utility of provider i is obtained by computing the Nash equilibrium. We get:

$$\frac{\partial U_i}{\partial p_i} = D - p_i d = 0, \;\; i = 1, 2. \tag{5.1}$$

This gives $p_1 = p_2 = D_0/(3d)$. The demand is now $D_0/3$, larger than in the cooperative case, and the revenue of each provider is $D_0^2/9$, less than before.

Next consider the competitive model and assume we install *side payments*: CP2 is requested to pay p_3 to ISP1 for "transit" costs. So, the revenues of the providers are:

$$\begin{aligned} U_1 &= [D_0 - (p_1 + p_2) \cdot d]\,(p_1 + p_3) \\ U_2 &= [D_0 - (p_1 + p_2) \cdot d]\,(p_2 - p_3) \end{aligned}$$

As the model so far is symmetric, we can in fact allow for negative value of p_3 which would model payment from the ISP1 to CP2 instead, *e.g.*, payment for copyright, as discussed below.

5.6 Discussion of side payments

At this point we render it asymmetric by assuming that p_3 is determined by ISP1 for the case $p_3 > 0$, *i.e.* additional transit revenue from the content provider in a "two sided" payment model to ISP1 [30, 31]. Then, unless $D = 0$ there is no optimal p_3: as it increases, so does U_1. Thus, at equilibrium necessarily $D = 0$, and the revenues of both service and content providers are 0. Hence p_1 and p_2 sum up to D_0/d. Then by decreasing p_1 slightly, the demand will become strictly positive, so ISP1 can increase its utility by U_1 without bound by choosing p_3 sufficiently large. Therefore, at equilibrium $p_1 = 0$ and $p_2 = D_0/d$. If $p_2 > p_3$ then by a slight decrease in p_2, U_2 strictly increases so this is not equilibrium. We conclude that at equilibrium, $p_3 \geq p_2$. To summarize, the set of equilibria is given by $\{p_1 = 0, p_2 = D_0/d$ and $p_3 \geq D_0/d\}$.

Thus by discriminating one provider over the other and letting it charge the other provider, both providers lose. Obviously the internauts do not gain anything either, as their demand is zero!

We have considered above side payment from the CP2 to ISP1. In practice, the side payment may go in the other direction. Indeed, there is a growing literature that argues that ISP1 has to pay to CP2. This conclusion is based on cooperative game theory (and in particular on Shapley values) which stipulates that if the presence of an economic actor A in a coalition creates revenue to another actor B, then actor A ought to be paid proportionally to the benefits that its presence in the coalition created. In our case, the CP2 creates a demand of users who need Internet access, and without the CP2, ISP1 would have less subscribers.

The use of Shapley value (and of a coalition game approach, rather than of a non-cooperative approach) has the advantage of achieving Pareto optimality. In particular this means that the total revenue for ISP1 and CP2 would be those computed under the cooperative approach.

Side payment to the CP2 from ISP1 may also represent payment to the copyright holders of the content being downloaded by the internauts. In particular, a new law is proposed in France, by a member of parliament of the governing party, to allow download of unauthorized copyright content and in return be charged *proportionally* to the volume of the download, with an average payment of about five euros per month. A similar law had been already proposed and rejected five years ago by the opposition in France. It suggested to apply a tax of about five euros on those who wish to be authorized to download copyrighted content. In contrast, the previously proposed laws received the support of the trade union of musicians in France. If these laws were accepted, the service providers would have been requested to collect the tax (that would be paid by the internauts as part of their subscription contract). Note that although $p_3 < 0$ in our model could represent these types of side payments, the copyright payments per user are actually not decision variables.

5.7 Revenue generated by advertising

We now go back to the basic collaborative model to consider the case where the CP2 has an additional source of revenue from advertisement that amounts to $p_4 D$. p_4 is assumed to be a constant. The total income of the providers is

$$\Pi = (D_0 - pd)(p + p_4) \tag{5.2}$$

Then

$$\frac{\partial \Pi}{\partial p} = D_0 - 2pd - dp_4 \tag{5.3}$$

Equating to zero, we obtain

$$p = \frac{D_0 - p_4 d}{2d} \tag{5.4}$$

The total demand is $(D_0 + p_4 d)/2$, and the total revenues at equilibrium are

$$U_t^{\max} = \frac{D_0^2 + 2\,d\,p_4\,D_0 + d^2\,p_4^2}{4\,d} \tag{5.5}$$

This result does not depend on the way the revenue from the internauts is split between the providers.

5.8 The case where $p_2 = 0$

In particular, the previous result covers the case where $p_2 = 0$, *i.e.*, the case where advertising is the only source of revenue for the content provider CP2. One may consider this to be the business model of the collective consisting of (i) BitTorrent permanent seeders and (ii) specialized torrent file resolvers (*e.g.*, Pirate Bay).

Note that BitTorrent permanent seeders may be *indifferent* to downloading to BitTorrent leecher clients (particularly during periods of time when the seeder workstations are not otherwise being used) because of flat-rate pricing for network access, *i.e.*, a flat-rate based on capacity without associated usage-based costs (not even as overages).

5.9 Best response

The utilities for the network access provider ISP1 and the content provider CP2 are, respectively,

$$U_1 = [D_0 - (p_1 + p_2) \cdot d] (p_1 + p_3) \tag{5.6}$$

and

$$U_2 = [D_0 - (p_1 + p_2) \cdot d] (p_2 - p_3 + p_4). \tag{5.7}$$

We first show that for any p_2, it is optimal for the ISP1 to choose $p_1 = 0$. First consider the problem of the best choice of p_1 and p_3 assuming the quantity $p_1 + p_3$ is constant; clearly, U_1 strictly decreases in p_1 so that a best response cannot have $p_1 > 0$.

Thus, if p_2 is not controlled (in particular if $p_2 = 0$ so that CP2's only revenue is from a third party and not directly from the users), then ISP1 would gain more by charging the CP2 than by charging the users. This is also consistent with the simple fact that $\partial U_1/\partial p_3 \geq \partial U_1/\partial p_1$.

5.10 Nash equilibrium

With $p_1 = 0$ and $p_3 \geq 0$, the utility of ISP1 is

$$U_1 = [D_0 - p_2 d] \, p_3 \tag{5.8}$$

Thus the condition on the best response of ISP1 for a given p_2 gives $p_2 = D_0/d$, *i.e.*, the demand is zero. On the other hand, for this p_2 to be a best response for U_2, $p_3 = p_2 + p_4$. We conclude that there is a unique Nash equilibrium given by $p_1 = 0$, $p_2 = D_0/d$, and $p_3 = D_0/d + p_4$.

5.11 Stackelberg equilibrium in network neutrality

Stackelberg equilibrium corresponds to another aspect of asymmetric competition, in which one competitor is a leader and the other a follower. Actions are no longer taken independently: here, first the leader takes an action, and then the follower reacts to this action.

Let's restrict to $p_3 \geq 0$.

We assume that the ISP1 is the leader. Given p_1 and p_3, U_2 is concave in p_2. So, a necessary and sufficient condition for p_2 to maximize this is

$$\frac{\partial U_2}{\partial p_2} = D_0 - d \cdot (p_2 - p_3 + p_4) - d \cdot (p_1 + p_2) = 0 \tag{5.9}$$

holds with equality for $p_2 > 0$. That is, to maximize U_2,

$$p_2 = \frac{1}{2}\Big(\frac{D_0}{d} + p_3 - p_1 - p_4\Big) > 0. \tag{5.10}$$

Substituting p_2 in U_1, we obtain:

$$\begin{aligned} U_1 &= [D_0 - (p_1 + p_2) \cdot d] (p_1 + p_3) \\ &= \frac{1}{2} [D_0 - 3p_1 d - p_3 d + p_4 d] (p_1 + p_3) \end{aligned}$$

We now compute the actions that maximize the utility U_1 which is concave in (p_1, p_3). We have

$$\frac{\partial U_1}{\partial p_1} = \frac{D_0 - 4dp_3 - 6dp_1 + dp_4}{2} \leq 0 \tag{5.11}$$

$$\frac{\partial U_1}{\partial p_3} = \frac{D_0 - 2dp_3 - 4dp_1 + dp_4}{2} \leq 0 \tag{5.12}$$

For $p_1 > 0$, (5.11) should hold as equality. Subtracting (5.11) from (5.12) we get $p_3 \leq -p_1$, and hence they are zero. This conclusion is in contradiction with our assumption $p_1 > 0$.

Assume that $p_1 = 0$ and $p_3 > 0$. Then U_1 is concave in p_3 and (5.12) holds with equality. Hence

$$p_3 = \frac{D_0}{2d} + \frac{p_4}{2} \tag{5.13}$$

maximizes U_1. Substituting in (5.10) we get

$$p_2 = \frac{1}{4}\Big(\frac{3D_0}{d} - p_4\Big) \tag{5.14}$$

We conclude that if $p_4 d < 3D_0$ Then the Nash equilibrium is $p_1 = 0$, and p_3 and p_2 are given, respectively, by (5.13) and (5.14).

Since we assume here that $p_2 \geq 0$, then in case $p_4 d \geq 3D_0$, we will have $p_2 = 0$ since this value maximizes (5.14).

5.12 Conclusions and on-going work

Using a simple, parsimonious model of linearly diminishing user/consumer demand as a function of price, we studied a game between collective players, the user ISP and content provider, under a variety of scenarios including: non-neutral two-sided transit pricing, copyright payments made by the ISP, the effects of flat-rate pricing, advertising revenue, cooperation, and leadership. In particular, we demonstrated under what conditions non-neutral transit pricing of content providers may result in revenue loss for all parties in play (*i.e.*, so that at least one player opts out of the game, where all players are necessary for positive outcome).

In on-going work, we are considering issues of non-monetary value and copyright. Moreover, we are including the users as active players. Finally, we are considering the effects of content-specific (not *application* neutral) pricing.

5.13 Exercise

Find all Stackelberg equilibria in the following coordination game given in Table 5.1. Are there equilibria in mixed strategies?

	action 1.a	action 1.b
action 2.i	2, 1	0, 0
action 2.ii	0, 0	1, 2

Table 5.1: A coordination game [Aumann]

Chapter 6

Further Notions of Equilia

6.1 The traffic assignment problem and Wardrop Equilibrium

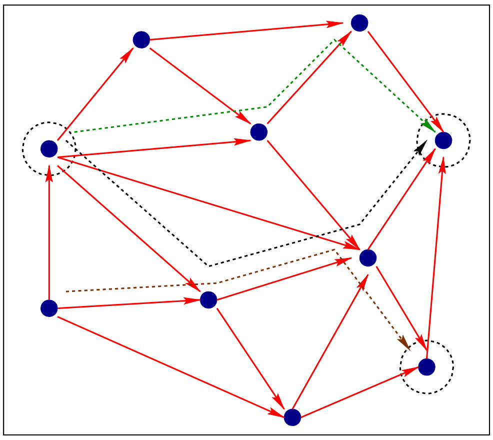


Figure 6.1: The assignment problem and competitive routing

Given the fundamental nature of equilibria in many large-scale systems, it is of no surprise that researchers studying transportation networks have been preoccupied with developing models that reproduce this equilibrium, as a function of network characteristics and user demand levels. Typically, transport equilibrium models consider vehicles to be the fundamental units seeking an equilibrium, or, in the case of public transport, the individual traveler. In both of these cases, since the number of users is generally very large, the Wardrop concept, that treats individual user contributions to the costs as infinitesimal, is preferred to the (in this respect, more general) Nash paradigm.

In the context of telecommunication networks, the Wardrop equilibrium is used most often to model the situation in which the routed entities are packets, and routing decisions are taken at the nodes of the networks (rather than by the users) so as to minimize the (per-packet) delay. In many actual networks, the routers at the nodes seek to minimize the per-packet delay in terms of the number of "hops," or nodes, to the destination. There are, however, situations in which it is more advantageous to work with actual delays as cost metrics, rather than the number of hops (see Bertsekas and Gallager 1987, Gupta and Kumar, 1997), and it is in these cases that the Wardrop equilibrium has been used to describe the resulting flow patterns. This is the case, for example, in ad-hoc networks in which both users as well as base stations are mobile, or where there are no base stations so that users are responsible to relay messages of other users. For these type of networks a Wardrop type equilibrium has been advocated in Bertsekas and Gallager, 1987.

Wardrop equilibria have also been used in telecommunication networks to model a large number of users that can determine individually their route and in which the routed object is a whole session, see Korilis and Orda (1999) (whose model includes in addition some side constraints on the quality of service).

A third context in which Wardrop equilibrium has been used outside of transportation is in distributed computer networks, in which the routed objects are jobs. An individual job can be processed in any of several interconnected nodes (computers) and the routing decision is taken so as to minimize its expected delay in the system (composed of both communication as well as processing delay). Much material on that application can be found in [32].

The definition of the steady state equilibrium of a traffic network was put forth by J.G. Wardrop in his 1952 treatise [33] which provided two different definitions of traffic assignment concepts. The first is commonly referred to as the Wardrop, or traffic equilibrium, principle and can be understood as a variant of Nash equilibrium for networks having a contimuum of players, where a single player is negligible. It states that "*The journey time on all the routes actually used are equal, and less than those which would be experienced by a single vehicle on any unused route.*"

The alternative social optimal framework is given by Wardrop's second principle, which states that "*The average journey time is a minimum.*" While some point to the economist Pigou, who stated analogous principles in his 1920 *Economics of Welfare*, as the rightful originator of these ideas, they had not been applied to networks, and transportation networks in particular, until Wardrop's seminal work.

The Nash equilibrium with finitely many decision makers is also relevant in the above road traffic framework. A decision maker could be a driver but could also correspond to a group of drivers for whom the trajectory is determined by some common decision omaker. When the decision makers in a game are discrete and finite in number, a Nash equilibrium can be achieved without the costs of all used routes being equal, contrary to Wardrop's equilibrium principle. In some cases, Wardrop's principle represents a limiting case of the Nash equilibrium principle, as the number of players becomes very large [34].

The second Wardrop principle, that of system optimality, assumes that congested networks can be globally optimized. While this can be true for a network which is entirely controlled by a single operator, it is not so with networks of road traffic or with disaggregated telecommunication networks.

The first Wardrop principle, stated above, can be expressed mathematically to state that the flow on every route r serving a commodity, or origin-destination (OD) pair, w, is either zero, or its cost is equal to the minimum cost on that OD pair. Along with the fact that the cost on any route serving an OD pair is at least as high as the minimum cost on that OD pair, and the satisfaction of demand for each OD pair, we obtain the following system:

$$h_{wr}(c_{wr} - \pi_w) = 0, r \in R_w, w \in W, \tag{6.1}$$

$$c_{wr} - \pi_w \geq 0, r \in R_w, w \in W, \tag{6.2}$$

$$\sum_{r \in R_w} h_{wr} = d_w, w \in W \tag{6.3}$$

where $h_{wr}$is the flow on route $r \in R_w$, the set of routes joining node pair $w \in W$,the set of origin-destination node-node pairs. The cost or delay on that route is c_{wr}, and π_w is the minimum cost on any route joining node pair w. The demand for service between the node pair w is denoted d_w. Define the graph as $G = (\mathcal{N}, \mathcal{A})$. Then, adding non-negativity restrictions $h_{wr} \geq 0$ and $\pi_w \geq 0$.

6.2 Evolutionary Stable Strategies

Consider a large population of players. Each individual needs occasionally to take some action. We focus on some (arbitrary) tagged individual. The actions of some M (possibly random number of) other individuals may interact with the action of the tagged individual (e.g. some other connections share a common bottleneck). In order to make use of the wealth of tools and theory developed in the biology literature, we shall restrict here (as they do), to interactions that are limited to pairwise, i.e. to $M = 1$. This will correspond to networks operating at light loads, such as sensor networks that need to track some rare events such as the arrival at the vicinity of a sensor of some tagged animal.

We define by $J(p, q)$ the expected payoff for our tagged individual if it uses a strategy p when meeting another individual who adopts the strategy q. This payoff is called “fitness" and strategies with larger fitness are expected to propagate faster in a population.

We assume that there are N pure strategies. A strategy of an individual is a probability distribution over the pure strategies. An equivalent interpretation of strategies is obtained by assuming that individuals choose pure strategies and then the probability distribution represents the fraction of individuals in the population that choose each strategy.

Suppose that the whole population uses a strategy q and that a small fraction ϵ (called “mutations") adopts another strategy p. Evolutionary forces are expected to select q against p if

$$J(q, \epsilon p + (1-\epsilon)q) > J(p, \epsilon p + (1-\epsilon)q) \tag{6.4}$$

A strategy q is said to be ESS if for every $p \neq q$ there exists some $\hat{\epsilon}_y > 0$ such that (6.4) holds for all $\epsilon \in (0, \hat{\epsilon}_y)$.

In fact, we expect that if for all $p \neq q$,

$$J(q, q) > J(p, q) \tag{6.5}$$

then the mutations fraction in the population will tend to decrease (as it has a lower reward, meaning a lower growth rate). The strategy q is then immune to mutations. If it does not but if still the following holds,

$$J(q, q) = J(p, q) \text{ and } J(q, p) > J(p, p) \quad \forall p \neq q \tag{6.6}$$

then a population using q are “weakly” immune against a mutation using p since if the mutant's population grows, then we shall frequently have individuals with strategy q competing with mutants; in such cases, the condition $J(q, p) > J(p, p)$ ensures that the growth rate of the original population exceeds that of the mutants. A strategy is ESS if and only if it satisfies (6.5) or (6.6), see [35, Proposition 2.1].

The conditions to be an ESS can be related to and interpreted in terms of Nash equilibrium in a matrix game. The situation in which an individual, say player 1, is faced with a member of a population in which a fraction p chooses strategy A is then translated to playing the matrix game against a second player who uses mixed strategies (randomizes) with probabilities p and $1-p$, resp. The central model that we shall use to investigate protocol evolution is introduced in the next subsection along with its matrix game representation.

Chapter 7

S-modular games

We have seen in several examples in previous Sections that equilibria best response strategies have either the "Join the Crowd" property or the the "Avoid the Crowd" property, which typically leads to threshold equilibria policies. These properties turn out to be also useful when we seek to obtain convergence to equilibria from a non-equilibrium initial point. These issues will be presented in this section within the framework of S-modular games due to Yao [36] who extends the notion of submodular games introduced by Topkis [37]. We then apply it to several queueing models.

S-modularity is then use to study a competitive model that describes the interaction between competing telecommunications service providers (SPs), their subscribers, and a network owner. Competition takes place in pricing as well as in terms of the quality of service (QoS) they offer. The subscribers' demand depends not only on the price and QoS of that SP but also upon those of its competitors. We establish conditions for existence and uniqueness of the equilibria, compute them explicitly and characterize their properties.

7.1 Model, definitions and assumptions

General model are developed in [37, 36] for games where the strategy space S_i of player i is a compact sublattice of $R^{\overline{m}}$. By sublattice we mean that it has the property that for any two elements x, y that are contained in S_i, also $\min(x, y)$ (denoted by $x \wedge y$) and $\max(x, y)$ (denoted by $x \vee y$) are contained there (by $max(x, y)$ we mean the componentwise max, and similarly with the min). We describe below the main results for the case that $\overline{m} = 1$.

Definition 3. *The utility f_i for player i is supermodular if and only if*

$$f_i(x \wedge y) + f_i(x \vee y) \geq f_i(x) + f_i(y).$$

It is submodular if the opposite inequality holds.

If f_i is twice differentiable then supermodularity is equivalent to

$$\frac{\partial^2 f_i(x)}{\partial x_1 \partial x_2} \geq 0.$$

Monotonicity of maximizers. The following important property was shown to hold in [37]. Let f be a supermodular function. Then the maximizer with respect to x_i is increasing in $x_j, j \neq i$.

More precisely, define the best response

$$BR_1^*(x_2) = argmax_{x_1} f(x_1, x_2);$$

if there are more than one argmax above we shall always limit ourselves to the smallest one (or always limit ourselves to the largest one). Then $x_2 \leq x_2'$ implies $BR_1^*(x_2) \leq BR_1^*(x_2')$. This monotonicity property holds also for non-independent policy sets such as (7.1), provided that they satisfy the ascending property (defined below).

Definition 4. (**Monotonicity of sublattices**) *Let A and B be sublattices. We say that $A \prec B$ if for any $a \in A$ and $b \in B$, $a \wedge b \in A$ and $a \vee b \in B$.*

Next, we introduce some properties on the policy spaces.
Consider two players. We allow S_i to depend on x_j

$$S_i = S_i(x_j), \quad i,j = 1,2, \quad i \neq j. \tag{7.1}$$

Monotonicity of policy sets We assume

$$x_j \leq x'_j \Longrightarrow S_i(x_j) \prec S_i(x'_j).$$

This is called the **Ascending Property**. We define similarly the **Descending Property**.

Lower semi continuity of policies We say that the point to set map $S_i(\cdot)$ is lower semi continuous if for any $x_j^k \to x_j^*$ and $x_i^* \in S_i(x_j^*)$ $(j \neq i)$, there exist $\{x_i^k\}$ s.t. $x_i^k \in S_i(x_j^k)$ for each k, and $x_i^k \to x_i^*$.

7.2 Existence of Equilibria and Round Robin algorithms

Consider an n-player game. Yao [36, Algorithm 1] and Topkis [37, algorithm I] consider a greedy round robin scheme where at some infinite strictly increasing sequence of time instants T_k, players update their strategies using each the best response to the strategies of the others. Player l updates at times T_k with $k = mn + l$, $m = 1, 2, 3, ...$

Assume lower semi-continuity and compactness of the strategy sets. Under these conditions, Supermodularity together with the ascending property imply monotone convergence of the payoffs to an equilibrium [36]. The monotonicity is in the same direction for all players: the sequences of strategies for each player either all increase or all decrease.

The same type of result is also obtained in [36, Thm. 2.3] with submodularity instead of supermodularity for the case of two players, where the ascending property is replaced by the descending property. The monotone convergence of the round robin policies still holds but it is in opposite directions: the sequence of responses of one player increases to his equilibrium strategy, while the ones of the other player decreases.

In both cases, there need not be a unique equilibrium.

Yao [36] further extends these results to cases of costs (or utilities) that are submodular in some components and supermodular in others. The notion of s-modularity is used to describe either submodularity or supermodularity. Another extension in [36] is to vector policies (i.e. a strategy of a player is in a sublattice of R^k).

Next we present several examples for games in queues where s-modularity can be used. The first two examples are due to Yao [36].

7.3 Example of supermodularity: queues in tandem

Consider a set of queues in tandem. Each queue has a server whose speed is controlled. The utility of each server rewards the throughput and penalizes the delay. Under appropriate conditions, it is then shown in [36] that the players have compatible incentives: if one speeds up, the other also want to speed up.

More precisely, consider two queues in tandem with i.i.d. exponentially distributed service times with parameters μ_i, $i = 1, 2$. Let $\mu_i \leq u$ for some constant u. Server one has an infinite source of input jobs There is an infinite buffer between server 1 and 2. **The throughput** is given by $\mu_1 \wedge \mu_2$.

The expected number of jobs in the buffer is given [36] by

$$\frac{\mu_1}{\mu_2 - \mu_1}$$

when $\mu_1 < \mu_2$, and is otherwize infinite.

Let

- $p_i(\mu_1 \wedge \mu_2)$ be the profit of server i,
- $c_i(\mu_i)$ be the operating cost,
- $g(\cdot)$ be the inventory cost.

The utilities of the players are defined as

$$f_1(\mu_1, \mu_2) := p_1(\mu_1 \wedge \mu_2) - c_1(\mu_1) - g\left(\frac{\mu_1}{\mu_2 - \mu_1}\right)$$

$$f_2(\mu_1, \mu_2) := p_2(\mu_1 \wedge \mu_2) - c_2(\mu_2) - g\left(\frac{\mu_1}{\mu_2 - \mu_1}\right).$$

The strategy spaces are given by

$$S_1(\mu_2) = \{\mu_1 : 0 \le \mu_1 \le \mu_2\},$$

$$S_2(\mu_1) = \{\mu_2 : \mu_1 \le \mu_2 \le u\}.$$

It is shown in [36] that if g is convex increasing then f_i are supermodular. So we can apply the results of the previous subsection, and obtain (1) the property of "joining the crowd" of the best response policies, (2) existence of an equilibrium, (3) convergence to equilibrium of some round robin dynamic update schemes.

7.4 Example of submodularity: flow control

We consider now an example for submodularity. There is a single queueing centre with two input streams with Poisson arrivals with rates λ_1 and λ_2. The rates of the streams are controlled by 2 players.

The queueing center consists of c servers and no buffers. Each server has one unit of service rate.

When all servers are occupied, an arrival is blocked and lost.

The blocking probability is given by the Erlang loss formula:

$$B(\lambda) = \frac{\lambda^c}{c!}\left[\sum_{k=0}^{c} \frac{\lambda^k}{k!}\right]^{-1}$$

where $\lambda = \lambda_1 + \lambda_2$.

Suppose user i maximizes

$$f_i = r_i(\lambda_i) - c_i(\lambda B(\lambda)).$$

c_i is assumed to be convex increasing. $\lambda B(\lambda)$ is the total loss rate.

Then it is easy to check [36] that f_i are submodular.

Two different settings can be assumed for the strategy sets. In the first, the available set for player i consists of $\lambda_i \le \overline{\lambda}$. Alternatively, we may consider that the strategy sets of the players depend on each other and the sum of input rates has to be bounded: $\lambda \le \overline{\lambda}$. Then S_i satisfy the descending property. We can thus apply again the results of the section 7.2.

7.5 A flow versus service control

Exponentially inter-arrivals as we used in previous examples are quite appealing to handle mathematically, and they can model sporadic arrivals, or alternatively, information packets that arrive one after the other but whose size can be approximated by an exponential random variable. In this example we consider, in contrast, a constant time T between arrivals of packets, which can be used for modeling ATM (Asynchronous Transfer Mode) networks in which information packet have a fixed size.

We consider a single node with a periodic arrival process, in which the first player controls the constant time period $T \in [\underline{T}, \overline{T}]$ between any two consecutive arrivals. We consider a single server with no buffer. The service time distribution is exponential with a parameter $\mu \in [\underline{\mu}, \overline{\mu}]$, which is controlled by the second player. If an arrival finds the server busy then it is lost. The loss probability is given by

$$P_l = \exp(-\mu T),$$

which is simply the probability that the random service time of (the previous) customer is greater than the constant T.

The transmission rate of packets is T^{-1}, but since a fraction P_l is lost then the goodput (the actual rate of packets that are transmitted successfully) is

$$G = \frac{1}{T}(1 - \exp(-\mu T)).$$

We assume that the utility of the first player is the goodput plus some function of the input rate T^{-1}. The server earns a reward that is also proportional to the goodput, and has some extra operation costs which is a function of the service rate μ. In other words,

$$J_1(T, \mu) = \frac{1}{T}(1 - \exp(-\mu T)) + f(T^{-1}), \qquad J_2(T, \mu) = \frac{1}{T}(1 - \exp(-\mu T)) + g(\mu).$$

We then have for $i = 1, 2$

$$\frac{\partial {J_i}^2}{\partial T \partial \mu} = -\mu \exp(-\mu T) \leq 0.$$

We conclude that the cost is submodular.

7.6 Modeling Network Equilibrium with Price and Quality-of-Service Characteristic

Competitive routing and equilibrium models for telecommunications networks generally assume a single characteristic through which an equilibrium is computed, such as delay [38], prices [39, ?], or even loss probabilities [?]. However, in order to take into account Quality of Service (QoS) on a network, it is generally necessary to incorporate into the model more than one parameter. A straightforward example is to include both prices and some measure of QoS. Other multi-criteria models may incorporate, for example, delay and reliability, the latter representing the QoS, price or delay and jitter, etc.

In this paper, we present a general model for computing a bi-criteria Nash equilibrium of several telecom providers. We shall then examine particular functional forms, and the resulting equilibrium properties, when the two parameters of each providers are prices and a measure of QoS.

Often, in research on competitive network routing and equilibrium models, the demands are assumed to be either given, as constants, or elastic but given by a function of a single parameter. In this paper we assume that the demand for the services of a given service provider can be described by a function that depends on the vectors of prices and qualities of service (QoS) of *all service providers.*

Our methodology is strongly inspired by [?] who studied a dynamic competitive inventory control model that includes both pricing as well as a quality parameter. The authors reduced the problem to a static game which is a special case of the general form of games that we shall study in our paper.

Section 7.7 provides the general model setting and some general theorems for existence and uniqueness of equilibria. Section 7.7.2 then presents the linear demand model and establishes some preliminary properties of the equilibrium for fairly general relation between the cost functions and QoS of the service providers, in the case where the QoS are fixed and competition occurs only through prices. The following two sections obtain stronger characterizations of the equilibria and analyze competition both in prices and in QoS. Two specific QoS measures are considered: delay (Section 7.8) and loss probabilities (Section 7.9). Section 7.9 includes several different possible ways of modeling loss probabilities including packet loss versus session rejection probability, and the influence of incorporating large-deviation scaling. We conclude with a number of promising avenues for extending these results.

Literature review

Whereas equilibria in transportation networks have been investigated for more than fifty years, this issue appeared much later in telecommunication networks, with the initial work on problems of competitive routing [?, 38], decentralized flow control [40, 41] and resource allocation [?]. One reason for the growing interest in competitive approaches to network management is the deregulation of the telecommunication monopolies. For surveys on competitive games in telecommunications, see [4, 3].

In our model, we do not take into account network topology, as in the above references, but rather model the total service proposed by each **SP** as a single entity. In other words, the price and QoS proposed by an **SP** will not depend on the physical source or destination, distance, etc. that underlies the request of each user. On the other hand, our model does take into account the quantities requested by users.

Telecommunications equilibria as a function of the price charged have been considered in great detail, as pricing has become a subject of intense research debate in the community. Numerous models exist, from utility maximizing approaches [?, ?] to auction-type approaches [?]. In these contributions, price and quantity requested have been the only parameters.

Another body of work has considered price as an endogeneous variable, which is determined as a function of the degree of saturation on the network. Typically in these approaches, the price is a *shadow price*, or Lagrange multiplier on inequality constraints, such as capacity restrictions. The reader is referred to [42, 43, 44, ?] for further details on those approaches. Finally, we mention work on defining equilibria on telecom networks as a function of other characteristics, such as loss probabilities, such as in [?].

In all of these approaches, however, a common point is that only one exogeneous characteristic gives rise to a competitive equilibrium. However, it is clear that, if one wishes to model QoS effects, more than one exogenous variable will need to be used. In this vein, J. Van Mieghem and P. Van Mieghem [?] consider the problem of pricing and scheduling of telecom services, in which the services are differentiated according to throughput, delay, and loss. While the authors consider fairness in scheduling, they do not compute equilibria or model competitive behavior. Park, Sitharam, and Chen [45] on the other hand model a “QoS provision game” in which QoS is defined by a vector of attributes. In their model, however, they do not provide existence or uniqueness results on the equilibria, as the particular user utility functions they employ do not satisfy sufficient concavity conditions. Rather, they provide results on identifying Nash equilibria, and Pareto, and system-optimal solutions.

7.7 Model and general properties

7.7.1 Model setting

Let us consider a game in which there are N service providers, **SP**, the set of which is denoted by $\mathcal{I} = \{1, 2, .., N\}$. Each provider has two parameters to set as regards the service it offers: $(\mathbf{p}, \mathbf{f}) \in \mathbb{R}_+^{N+N}$. (We use bold face to denote a vector, e.g. $\mathbf{p}$ and denote its ith component by p_i.) Although the two parameters can be quite general, we shall regard the first as prices (p_i is the price that **SP** i charges his subscribers per unit demand) and the second as some QoS measure (delay, loss or rejection probability, etc). Then, **SP** i experiences a demand for its service, $D_i : \mathbb{R}_+^{N+N} \mapsto \mathbb{R}_+$ which depends not only on its own parameters, p_i and f_i, but also on the prices and QoS offered by its competitors; that is, each demand function D_i depends upon the entire price vector $\mathbf{p} = (p_1, p_2, .., p_N)$ and the entire QoS vector $\mathbf{f} = (f_1, ..., f_N)$. The demand functions themselves can take on a number of functional forms, each with its own consequences upon the resulting equilibrium. In this work, however, we shall consider only a linear demand model, which will be defined further in the following section.

The utility functions of the service providers are given by $U : \mathbb{R}_+^{N+N} \mapsto \mathbb{R}_+$. Finally, we must define the strategy space of the N service providers. We shall suppose that all service providers' strategies are defined by the following orthogonal constraints; the strategy space, $\mathcal{R}_i$, of provider i is given by the subset of $\mathbb{R}^2$:

$$\mathcal{R}_i = \{(p_i, f_i) : \ 0 \leq \hat{g}_i(f_i) \leq p_i \leq p^{\max}; \quad 0 \leq f_i^{\min} \leq f_i \leq f_i^{\max}\}, \tag{7.2}$$

where $p_i^{\min} = \hat{g}_i(f_i)$. (The dependence of $p_i^{\min}$ on f_i through the function $\hat{g}_i$ will allow us to eliminate explicitly policies that will result in negative profits for a service provider). The fact that there are maximum prices reflects the fact that beyond some reasonable price, the demand will be zero (what ever the prices of competitors and quality of services are). In addition, $\hat{g}_i$ is assumed convex so that the strategy space of player i is convex, for all $i \in \mathcal{I}$. (This will be needed in Theorem 4). We further define:

$$\begin{aligned}
\mathcal{R}_{pi}(f_i) &= \{p_i : \ 0 \leq \hat{g}_i(f_i) \leq p_i \leq p^{\max}\}, && (7.3)\\
\mathcal{R}_{fi} &= \{f_i : \ 0 \leq f_i^{\min} \leq f_i \leq f_i^{\max}\}, && (7.4)\\
\mathcal{R} &= \mathcal{R}_1 \times \cdots \mathcal{R}_N && (7.5)\\
\mathcal{R}_p(\mathbf{f}) &= \mathcal{R}_{p1}(\mathbf{f}) \times \cdots \mathcal{R}_{pN}(\mathbf{f}) && (7.6)\\
\mathcal{R}_f &= \mathcal{R}_{f1} \times \cdots \mathcal{R}_{fN}. && (7.7)
\end{aligned}$$

Along with the model of the service providers, we include a model of the (single) network owner, who may charge each service provider, i, a cost per unit of bandwidth requested as a function of the two criteria it requires, p_i and f_i. Then, the total price charged to the service provider will be a function of that per-unit price and of the demand that the provider i experiences. That is, given the aggregate demand $d_i = D_i(\mathbf{p}, \mathbf{f})$ of **SP** i, we associate with **SP** i a cost function given by $R_i(p_i, f_i, d_i)$, representing the fee that i pays to the network owner.

We next define the Nash equilibria that we are interested in computing. The first involves each service provider setting a single parameter, as a function of those of all other providers, and the second involves simultaneous setting of both parameters by all providers, as a function of those of all other providers. As already mentioned, our approach is inspired by [**?**] who studied a dynamic competitive inventory control model that include both pricing as well as a quality parameter.

Definition 5. [Single-parameter Nash equilibrium] *Let $U_i(\mathbf{p}, \mathbf{f})$ be the net revenue*[1] *of service provider i, when the vector of prices set by all the providers is given by $\mathbf{p}$, and the vector of Quality of Service parameters, $\mathbf{f}$, of all the providers is fixed at some predetermined point, $\hat{\mathbf{f}}$. Then, a single-parameter Nash equilibrium in $\mathbf{p}$ at $\hat{\mathbf{f}}$ is the vector $\mathbf{p}^*$ that solves the following system for all $i \in \mathcal{I}$.*

[1] The net revenue can be expressed as the income minus the cost R_i. Examples will be given in later sections.

$$U_i(\mathbf{p}^*, \hat{\mathbf{f}}) = \max_{p_i \in \mathcal{R}_{pi}(\hat{\mathbf{f}})} U_i(p_1^*, ..., p_{i-1}^*, p_i, p_{i+1}^*, .., p_N^*, \hat{\mathbf{f}}). \tag{7.8}$$

This equilibrium corresponds to a price equilibrium for a fixed QoS vector $\hat{\mathbf{f}}$.

Definition 6. [Two-parameter Nash equilibrium] *Let* $U_i(\mathbf{p}, \mathbf{f})$ *be the net revenue of service provider* i*, when the pair of vectors of parameters set simultaneously by all the providers is given by* $(\mathbf{p}, \mathbf{f})$*. Then, a two-parameter Nash equilibrium in* $(\mathbf{p}, \mathbf{f})$ *is the vector couple* $(\mathbf{p}^*, \mathbf{f}^*)$ *that solves the following system for all* $i \in \mathcal{I}$*.*

$$U_i(\mathbf{p}^*, \mathbf{f}^*) = \max_{(p_i, f_i) \in \mathcal{R}_i} U_i(p_1^*, ..., p_{i-1}^*, p_i, p_{i+1}^*, .., p_N^*, f_1^*, ..., f_{i-1}^*, f_i, f_{i+1}^*, .., f_N^*). \tag{7.9}$$

7.7.2 General demand model

We shall suppose that the average demand, $D_i(\mathbf{p}, \mathbf{f})$, is linear in all prices and QoS levels. For a particular **SP** i, the demand for its service, D_i should be decreasing in the price it charges, p_i, but increasing in the prices charged by its competitors, p_j, $j \neq i$. The analogous relationship holds in service quality, but in that case, D_i is increasing in f_i and decreasing in f_j, for $j \neq i$. Then, we may write

$$D_i(\mathbf{p}, \mathbf{f}) = a_i - b_i p_i + \sum_{j \in \mathcal{I},\, j \neq i} c_{ij} p_j + \beta_i f_i - \sum_{j \in \mathcal{I},\, j \neq i} \gamma_{ij} f_j, \tag{7.10}$$

with b_i, c_{ij}, β_i and γ_{ij} positive constants.

We shall make the following blanket assumptions, which will be in effect throughout this study:

Assumption 1. *Without loss of generality, we shall assume that the demand function of each service provider satisfies, for all* $i \in \mathcal{I}$,

$$D_i(p^{\max} - p_i^{\min}, \mathbf{0}_{-i}, \mathbf{f}) \leq 0, ; \quad \forall \mathbf{f} \in \mathcal{R}_f, \tag{7.11}$$

where, as before, $p_i^{\min} = \hat{g}_i(f_i)$*, and* $\mathbf{0_{-i}}$ *is the* $N-1$*-dimensional null vector.*

The above assumption should be understood simply as imposing some relation between the parameters that define the demand functions. It could seem an realistic assumption, since it requires the demand to possibly be negative. However, we note that the nonpossitivity of the demand function is required to hold for values of the prices that are not necessarily within the set of feasible strategies $\mathcal{R}_i$'s. Therefore the assumption could be understood as requiring that if we extrapolated linearly the demand out of the feasible region and allowed the prices of all other service providers to be zero, then we would get a nonpositive demand.

We would expect, in practice, that an even stronger assumption would likely be satisfied, namely that $D_i(p_i, \mathbf{0}_{-i}, \mathbf{f}) \leq 0, ; \quad \forall \mathbf{f} \in \mathcal{R}_f; \quad \forall \mathbf{p_i} \in \mathcal{R}_{pi}(f_i)$. We require, for technical purposes, however, only this weaker assumption, that the demand for each provider's services be non-positive when competitor's prices are null, and his own prices are given by $p^{\max} - p_i^{\min} = p^{\max} - \hat{g}_i(f_i)$.

Assumption 2. *The constants* b *and* c *satisfy:*

$$b_i > \sum_{j \neq i} c_{ij}, \quad i \in \mathcal{I}. \tag{7.12}$$

Moreover, assume that D_i *is non-negative over the strategy space.*

Sufficient condition for Assumption 1 : Let Assumption A.6 hold. In view of (15.3), if we choose $p^{\max}$ large enough then Assumption 1 will hold.

Assumption A.6 will be needed to ensure the uniqueness of the resulting equilibrium. It is furthermore a reasonable condition, in that Assumption A.6 implies that the influence of an **SP**'s price is significantly greater on its observed demand than the prices of its competitors. This condition could then take into account the presence of customer loyalties and/or imperfect knowledge of competitors' prices.

7.7.3 General QoS model

Along with the model of the service providers, the (single) network owner charges each service provider, i, a cost per unit of bandwidth requested. We assume that the amount of bandwidth μ_i requested by **SP** i depends on the demand it experiences and on the QoS it wishes to offer (the higher the demand and the better the QoS, the higher μ_i will be). For example, if the QoS is linear in the standard Kleinrock delay function (obtained from an M/M/1 queuing model), i.e. $f_i = c - 1/(\mu_i - d_i)$, then $\mu_i(f_i, d_i) = d_i + 1/(c - f_i)$.

Thus, given the aggregate demand $d_i = D_i(p_i, f_i)$, the fee paid by **SP** i can be written as

$$R_i = v_i \mu_i(f_i, d_i)$$

where $\mu_i(f_i, d_i) \in [0, B]$ is again the bandwidth required in order to guarantee the QoS, f_i, at a demand level of d_i for **SP** i, and v_i is the price to be paid per unit of bandwidth.

7.7.4 Utility model

The revenue of **SP** i is given by $p_i D_i(\mathbf{p}, \mathbf{f})$, whereas the profit (or net revenue) of **SP** i is the difference between the revenue and the fee it pays to the network owner,

$$U_i(\mathbf{p}, \mathbf{f}) = p_i D_i(\mathbf{p}, \mathbf{f}) - R_i(f_i, d_i). \tag{7.13}$$

In the sequel, we shall show the uniqueness of Nash equilibrium for the price game (7.8) in which the capacity μ_i required by user i has the following form:

$$\mu_i(f_i, d_i) = d_i g_i(f_i) + h_i(f_i), \tag{7.14}$$

where g_i and h_i are positive functions.

Hence, the profit function becomes

$$U_i(\mathbf{p}, \mathbf{f}) = D_i(\mathbf{p}, \mathbf{f})(p_i - v_i g_i(f_i)) - v_i h(f_i).$$

In the single-parameter Nash game, the strategy space of each **SP** i is given by (7.2) and in the two-parameter game, the strategy space is given by (7.3) where $p^{\min} = \hat{g}_i(f_i) = v_i g_i(f_i)$. Note that the sets $\mathcal{R}_i$, and $\mathcal{R}_{pi}$, $i \in \mathcal{I}$ are convex when g_i is a convex function.

7.7.5 Properties of the general model

The question is then, under general assumptions on the model functions and data, when can we ensure the existence and uniqueness of the resulting equilibrium across service providers. To this end, we make the following assumptions that concern the properties of the utility functions.

Assumption 3. *Without loss of generality, express the net revenue (or utility) of service provider i as $U_i = \hat{U}_i(\mathbf{p}, \mathbf{f}, d_i) = U_i(\mathbf{p}, \mathbf{f})$, $i \in \mathcal{I}$ where, as before, $d_i = D_i(\mathbf{p}, \mathbf{f})$. Assume that U_i is upper semi-continuous in $(\mathbf{p}, \mathbf{f}) \in \mathcal{R}$, for all $i \in \mathcal{I}$.*

Under Assumption 3, one has the following general result on the existence of a Nash equilibrium among the N service providers.

Theorem 4. [Existence] *Under the Assumption 3 and the definition of the strategy spaces $\mathcal{R} = \mathcal{R}_i \times \cdots \mathcal{R}_N$, there exists at least one single-parameter (resp., two-parameter) Nash equilibrium of the service providers' game.*

Proof: The strategy space for each $i \in \mathcal{I}$ is convex and compact. The utility functions are assumed u.s.c in both arguments. The existence of a Nash point follows (see e.g. [46]). □

We first consider the single-parameter price game (7.8) which arises when the vector of quality of service levels of all providers is fixed. Then, we shall consider the two-parameter price and QoS game.

The following theorem is a generalization of [?, Thrm. 5] that established uniqueness of an equilibrium for a similar setting but when $h_i(.)$ are all zero and for some specific functions $g_i(.)$. By allowing for general g_i and h_i we shall be able to handle in the next sections a large number of quality of service functions that are typical in networks.

Theorem 5. [Uniqueness of Price-based Nash equilibrium] *Consider the price game (7.8) which arises when the vector of quality of service levels* $\mathbf{f}$ *is fixed, and assume that* $\mathcal{R}_i(\mathbf{f})$ *are nonempty,* $i \in \mathcal{I}$. *Then*

1. *The price game has a unique equilibrium* $\mathbf{p}^*(\mathbf{f})$, *with*

$$\mathbf{p}^*(\mathbf{f}) = (I - T)^{-1}\Gamma^{-1}q \tag{7.15}$$

where $\Gamma = diag(2b_1, ..., 2b_N)$, $T_{ii} = 0$, $T_{ij} = \frac{c_{ij}}{2b_j}$ *for* $i \neq j$ *and* $q_i = b_i v_i g_i(f_i) + a_i + \beta_i f_i - \sum_{j \neq i} \gamma_{ij} f_j$.

2. *The equilibrium prices* p_i^* *are increasing in each of the unit prices* v_j.

3. *The equilibrium prices* p_i^* *are decreasing in* f_j, $j \in \mathcal{I}$ *when*

$$g_j'(f_j) < \frac{\sum_{l \in \mathcal{I}} (L^{-1})_{il}\gamma_{jl} - (L^{-1})_{ij}\beta_j}{(L^{-1})_{ij} b_j v_j}$$

and increasing otherwise.

Proof: The existence of a Nash equilibrium of the price game (7.8) was demonstrated in Theorem 4. Therefore, we show the uniqueness of the equilibrium point.

Since $\mathbf{p}^*(\mathbf{f})$ is a Nash equilibrium, it follows that p_i^* maximizes the function $U_i(\mathbf{p}_i, \mathbf{p}^*_{-i}, \mathbf{f})$, where $\mathbf{p}_{-i} = (p_1, .., p_{i-1}, p_{i+1}, .., p_N)$. Moreover , we have

$$\begin{aligned}
\lim_{p_i \to v_i g_i(f_i)} \frac{\partial U_i(p_i, p^*_{-i}, \mathbf{f})}{\partial p_i} &= d_i(v_i g_i(f_i), p^*_{-i}, \mathbf{f}) \geq 0 && (7.16)\\
\lim_{p_i \to p^{max}} \frac{\partial U_i(p_i, p^*_{-i}, \mathbf{f})}{\partial p_i} &= -b_i p^{max} + b_i v_i g_i(f_i) + D_i(p^{\max}, \mathbf{p}^*_{-i}, \mathbf{f}) \\
&= a_i - 2b_i p^{\max} + \sum_{j \neq i} c_{ij} p_j^* + b_i v_i g_i(f_i) + \beta_i f_i - \sum_{j \neq i} \gamma_{ij} f_j \\
&\leq a_i - 2b_i p^{\max} + p^{\max} \sum_{j \neq i} c_{ij} + b_i g_i(f_i) + \beta_i f_i - \sum_{j \neq i} \gamma_{ij} f_j && (7.17)\\
&< a_i - b_i p^{\max} + b_i v_i g_i(f_i) + \beta_i f_i - \sum_{j \neq i} \gamma_{ij} f_j && (7.18)\\
&= d_i(p^{\max} - v_i g_i(f_i), \mathbf{0}_{-i}, \mathbf{f}) \\
&\leq 0 && (7.19)
\end{aligned}$$

where the transition (7.17)-(7.18) is due to the Assumption A.6 and the last transition is due to Assumption (1). Thus, p_i^* is unique solution of the optimality conditions:

$$0 = \frac{\partial U_i(\mathbf{p}, \mathbf{f})}{\partial p_i} = -b_i(p_i - v_i g_i(f_i)) + d_i(\mathbf{p}, \mathbf{f}). \tag{7.20}$$

Then, $\mathbf{p}^*$ is a solution to the following linear system

$$Lp = q \tag{7.21}$$

where $L_{ii} = 2b_i$, $L_{ij} = -c_{ij}$ for $i \neq j$ and $q_i = b_i v_i g_i(f_i) + a_i + \beta_i f_i - \sum_{j \neq i} \gamma_{ij} f_j$.
To show the uniqueness of Nash equilibrium, it suffices to show that the linearsystem (7.21)

admits a unique solution. We observe that the matrix L can be written as: $L = \Gamma(I-T)$ with $\Gamma = diag(2b_1, ..., 2b_N)$, and

$$T = \begin{pmatrix} 0 & \vdots & \frac{c_{1N}}{2b_1} \\ & \ddots & \\ \frac{c_{N1}}{2b_N} & \vdots & 0 \end{pmatrix}.$$

The matrix T is substochastic since $\frac{\sum_{j\neq i} c_{ij}}{2b_i} \leq \frac{b_i}{2b_i} < 1$. Thus L^{-1} exists and is equal to $(I-T)^{-1}\Gamma^{-1}$ and

$$p^*(\mathbf{f}) = L^{-1}q = (I-T)^{-1}\Gamma^{-1}q.$$

Writing explicitly one term from the equilibrium price vector, $p_i^*(\mathbf{f})$, and rearranging terms to identify dependencies on f_i and f_j independently, $j \in \mathcal{I}$, $j \neq i$, we have that

$$\begin{aligned} p_i^*(\mathbf{f}) &= (L^{-1})_{ii}(a_i + (b_i v_i g_i(f_i)) + f_i((L^{-1})_{ii}\beta_i - \sum_{j\neq i}(L^{-1})_{ij}\gamma_{ji}) + \\ & \sum_{j\neq i}(L^{-1})_{ij}(a_j + b_j v_j g_j(f_j) + \beta_j f_j) - \sum_{j\in\mathcal{I}}\sum_{l\neq i}(L^{-1})_{ij}\gamma_{jl} f_l \end{aligned} \tag{7.22}$$

Then, taking partial derivatives, we obtain:

$$\frac{\partial p_i^*(\mathbf{f})}{\partial f_j} = (L^{-1})_{ij} b_j v_j g_j'(f_j) + (L^{-1})_{ij}\beta_j - \sum_{l\in\mathcal{I}}(L^{-1})_{il}\gamma_{jl}, \tag{7.23}$$

which is negative when

$$g_j'(f_j) \quad < \quad \frac{\sum_{l\in\mathcal{I}}(L^{-1})_{il}\gamma_{jl} - (L^{-1})_{ij}\beta_j}{(L^{-1})_{ij} b_j v_j}. \tag{7.24}$$

□

Next we make the following assumptions on the functions g_i and h_i appearing in Definition (7.14) of μ_i. This will guarantee the uniqueness of the equilibrium $(\mathbf{p}^*, \mathbf{f}^*)$ in the price-QoS game (7.9) which we shall compute. Several examples of this structure will be given in Sections 7.8 and 7.9.

Assumption 4. *Let $h_i(.) = 0$. Moreover, suppose that $g(.) \in C^2$ is convex increasing, and its derivative satisfies for all $i \in \mathcal{I}$*

$$\lim_{f_i \to f_i^{\max}} g_i'(f_i) = +\infty.$$

This assumption will be satisfied in most of the queuing systems we study. The following result establishes the uniqueness of Nash equilibrium in which prices and QoS levels are set by all players simultaneously.

Theorem 6. [Uniqueness of Price and QoS-based Nash equilibrium] *Ket Assumption* **??** *hold. The price-QoS problem (7.9) has a unique two-parameter Nash equilibrium $(\mathbf{p}^*, \mathbf{f}^*)$, with f_i^* the unique solution to the following equations, for every $i \in \mathcal{I}$,*

$$\begin{aligned} g_i'(f_i^*) &= \frac{\beta_i}{v_i b_i} \text{ if } \frac{\beta_i}{v_i b_i} \geq g_i'(f_i^{\min}) \\ f_i^* &= 0 \text{ if } \frac{\beta_i}{v_i b_i} < g_i'(f_i^{\min}) \end{aligned}$$

Proof: Let

$$\tilde{U}_i(\mathbf{p}, \mathbf{f}) = \log[p_i - v_i g_i(f_i)] + \log[D_i(\mathbf{p}, \mathbf{f})].$$

Note that a vector $(\mathbf{p}^*, \mathbf{f}^*)$ is an equilibrium in the original utilities if and only if it is an equilibrium in terms of their logarithms, i.e. of the functions $\tilde{U}_i$.

Since g_i is convex by assumption, U_i is concave in $(\mathbf{p}, \mathbf{f})$. Under Assumption 4, we have that

$$\lim_{f_i \to f_i^{\max}} \frac{\partial \tilde{U}_i(\mathbf{p}, \mathbf{f})}{\partial f_i} = \lim_{f_i \to f_i^{\max}} \frac{-v_i g'(f_i)}{p_i - v_i g_i(f_i)} + \frac{\beta_i}{D_i(\mathbf{p}^*, \mathbf{f})} = -\infty. \tag{7.25}$$

This implies there will not be a Nash equilibrium at the constraint boundary where $f_i = f_i^{\max}$, $i \in \mathcal{I}$. To show the uniqueness of the Nash solution, we must thus examine the interior of the feasible region with respect to the upper bound on the f_i and show that there is precisely one point, $(\mathbf{p}^*, \mathbf{f}^*)$, which solves simultaneously the utility maximization problems of each provider, $i \in \mathcal{I}$. Note from (7.25) that the lower limit, $\lim_{f_i \to f_i^{\min}} \partial \tilde{U}_i(\mathbf{p}, \mathbf{f})/\partial f_i$ may be either positive or negative, and as such, the Nash equilibrium may arise at a point for which $f_i^* = f_i^{\min}$ for some $i \in \mathcal{I}$.

On the other hand, we do not need Lagrange multipliers for the constraints on p_i since p_i^* is obtained at a point where the partial derivative (7.20) is zero, possibly at the lower boundary, (7.16) and (7.19).

That is, it is necessary to consider the following Kuhn-Tucker conditions in which a Lagrange multiplier, α_i represents the lower bound constraint on f_i for each i. In other words, $(\mathbf{p}^*, \mathbf{f}^*)$ is a Nash equilibrium if for every $i \in \mathcal{I}$, there exist non-negative Lagrange multipliers α^i such that

$$\begin{aligned}
0 &= \frac{\partial \tilde{U}_i(p, \mathbf{f})}{\partial p_i} = \frac{1}{(p_i - v_i g_i(f_i))} + \frac{-b_i}{D_i(p^*, \mathbf{f})} && (7.26)\\
-\alpha^i &= \frac{\partial \tilde{U}_i(p, \mathbf{f})}{\partial f_i} = \frac{-v_i g_i'(f_i)}{p_i - v_i g_i(f_i)} + \frac{\beta_i}{D_i(p^*, \mathbf{f})} && (7.27)\\
\alpha^i(f_i - f_i^{\min}) &= 0, \quad \alpha^i \geq 0 && (7.28)
\end{aligned}$$

By multiplying (7.26) by $1/b_i$ and (7.27) by $1/\beta_i$ and substituting the resulting equations, we obtain

$$\frac{\alpha^i}{\beta_i} = \frac{v_i g_i'(f_i)/\beta_i - 1/b_i}{p_i - v_i g_i(f_i)}, \quad \text{and } \alpha_i(f_i - f_i^{\min}) = 0 \tag{7.29}$$

It follows that

$$\begin{aligned}
g_i'(f_i^*) &= \frac{\beta_i}{v_i b_i} \text{ if } \frac{\beta_i}{v_i b_i} \geq g_i'(f_i^{\min})\\
f_i^* &= f_i^{\min} \text{ if } \frac{\beta_i}{v_i b_i} < g_i'(f_i^{\min})
\end{aligned}$$

Fixing $f = f^*$, we obtain p^* through (7.15). Thus $(\mathbf{p}^*, \mathbf{f}^*)$ is the unique Nash equilibrium. □

Next, we consider two applications of this framework, making use of two different QoS measures.

7.8 Case study I: expected delay as QoS

We shall suppose in this section that the measure defining the QoS, f, corresponds to some function of the expected delay. We shall use the Kleinrock delay function[2] which is a common delay function used in networking games [38]. Instead of Minimizing delay, we consider the maximization of the reciprocal of its square-root (in that way the QoS f_i indeed increases as the delays decrease (this is similar to [?]):

$$f_i = \frac{1}{\sqrt{Delay}} = \sqrt{\mu_i - d_i} \tag{7.30}$$

[2]This function corresponds to a queuing delay in an $M/M/1$ queue with first-in-first-out discipline or to the more general $M/G/1$ queue under processor sharing delay.

Thus $\mu_i(f_i, d_i) = d_i g(f_i) + h_i(f_i)$ where $g_i(f_i) = 1$ is indeed convex and where $h_i(f_i) = f_i^2$. The cost that **SP** i pays the network owner is $v_i[(f_i^2) + d_i]$. Then, the net revenue of **SP** i that it wishes to maximize is then

$$U_i(\mathbf{p}, \mathbf{f}) = D_i(\mathbf{p}, \mathbf{f})(p_i - v_i) - v_i f_i^2. \tag{7.31}$$

We observe that this profit function is special case of the one defined in (7.14), by taking $g_i(f_i) = 1$ and $h_i(f_i) = f_i^2$.

We shall make the following assumption:

Assumption 5. *The constants b, v, and β satisfy:*

$$2b_i v_i > \beta_i^2, \quad i \in \mathcal{I} \tag{7.32}$$

Remark 6. *Note that, for price-sensitive demands, i, this assumption is likely to hold, since in that case, $b_i >> \beta_i$. For relatively price-*in*-sensitive demands, who value above all QoS, Assumption 5 may still hold if the unit price for capacity charged by the network owner to* **SP** *i, v_i* **is high.**

We now proceed with the analysis of the price-QoS game (7.9) (two-parameter) in which prices and services levels are set by all **SP**s simultaneously.

Theorem 7. *Let Assumption 5 hold. The price-QoS game (7.9) with QoS, $\mathbf{f}$ given by (7.30) has a unique Nash equilibrium $\mathbf{p}^*, \mathbf{f}^*$), with p_i^* increasing in each v_j, $j \in \mathcal{I}$.*

Proof: Assumption 3 is satisfied by (7.9) with (7.30) and (7.31). Hence, by Theorem 4, there exists a Nash equilibrium.

We next show that the Nash equilibrium is unique. We have seen, in the previous section, that the partial derivatives of the profit functions with respect to price variables p_i of each providers i will be zero. In this section, the upper and lower bounds on the quality of service variables, f_i may or may not be active, and so we take a Lagrangian relaxation of those constraints and examine the Kuhn Tucker conditions of the resulting Lagrangian.

To show the uniqueness of the Nash equilibrium solution, we show first that the profit functions U_i are jointly strictly concave in (p_i, f_i) for all providers, $i \in \mathcal{I}$, so that the solution to system defined by the Kuhn Tucker conditions is uniquely defined.

To this end, note first that

$$\begin{aligned}
\frac{\partial U_i(p, \mathbf{f})}{\partial p_i} &= -b_i(p_i - v_i) + D_i(p^*, \mathbf{f}) \text{ and } \frac{\partial U_i(p, \mathbf{f})}{\partial f_i} = \beta_i(p_i - v_i) - 2v_i f_i \\
\frac{\partial^2 U_i(p, \mathbf{f})}{\partial p_i^2} &= -2b_i < 0, \ \frac{\partial^2 U_i(p, \mathbf{f})}{\partial p_i \partial f_i} = \beta_i, \text{ and } \frac{\partial^2 U_i(p, \mathbf{f})}{\partial f_i^2} = -2v_i < 0.
\end{aligned}$$

Since $\frac{\partial^2 U_i}{\partial f_i^2} = -2v_i < 0$ and $\frac{\partial^2 U_i}{\partial p_i^2} = -2b_i < 0$, it suffices to show that the determinant of the Hessian matrix H is positive, where

$$H = \begin{pmatrix} \frac{\partial^2 U_i}{\partial p_i^2} & \frac{\partial^2 U_i}{\partial p_i \partial f_i} \\ \frac{\partial^2 U_i}{\partial f_i \partial p_i} & \frac{\partial^2 U_i}{\partial f_i^2} \end{pmatrix}$$

We have

$$\det H[U_i(\mathbf{p}, \mathbf{f})] = \frac{\partial^2 U_i}{\partial f_i^2} \frac{\partial^2 U_i}{\partial p_i^2} - \left(\frac{\partial^2 U_i}{\partial p_i \partial f_i} \right)^2 = 4b_i v_i - \beta_i^2 > 0$$

The latter inequality follows from Assumption 5.

We note that any solution to the following Kuhn-Tucker conditions is a Nash equilibrium: For every $i \in \mathcal{I}$, there exist non-negative Lagrange multipliers α^i and λ^i such that

$$\begin{aligned} 0 &= \frac{\partial U_i(\mathbf{p},\mathbf{f})}{\partial p_i} = -b_i(p_i - v_i) + D_i(\mathbf{p}^*,\mathbf{f}) && (7.33)\\ \alpha^i - \lambda^i &= \frac{\partial U_i(\mathbf{p},\mathbf{f})}{\partial f_i} = \beta_i(p_i - v_i) - 2v_i f_i && (7.34)\\ \alpha^i(f_i - f_i^{\min}) &= 0,\ \ \lambda^i(f_i^{\max} - f_i) = 0 && (7.35)\end{aligned}$$

where α^i and λ^i are the Lagrange multipliers corresponding to the min and max constraints on the values of f_i. As before, we do not need Lagrange multipliers for the constraints on p_i since p_i^* is obtained at a point where the partial derivative (7.33) is zero, possibly at the lower boundary, (7.16) and (7.19).

It follows from (7.34) that

$$\begin{aligned} \text{if } \lambda^i = 0 \text{ and } \alpha^i = 0,\ f_i^* &= \frac{\beta_i}{2v_i}(p_i - v_i) && (7.36)\\ \text{if } \alpha^i > 0,\ f_i^* &= f_i^{\min} && (7.37)\\ \text{if } \lambda^i > 0,\ f_i^* &= f_i^{\max} && (7.38)\end{aligned}$$

Hence, it follows that $\mathbf{p}^*$ is solution of the following linear system

$$Ap = r \qquad (7.39)$$

where

$$\begin{aligned} A_{ii} &= 2b_i - \delta_{\{\lambda^i - \alpha^i = 0\}}\frac{\beta_i^2}{2v_i},\\ A_{ij} &= -c_{ij} + \delta_{\{\lambda^j - \alpha^i = 0\}}\gamma_{ij}\frac{\beta_j}{2v_j},\\ R_i &= b_i v_i + a_i + \delta_{\{\lambda^i - \alpha^i = 0\}}\frac{\beta_i^2}{2} - \sum_{j\neq i}\delta_{\{\lambda^j - \alpha^j = 0\}}\gamma_{ij}\frac{\beta_j}{2} + \beta_i(\delta_{\{\lambda^i>0\}}f_i^{\max} + \delta_{\{\alpha^i>0\}}f_i^{\min})\\ &\quad - \sum_{j\neq i}\gamma_{ij}(\delta_{\{\lambda^i>0\}}f_j^{\max} + \delta_{\{\alpha^j>0\}}f_i^{\min}).\end{aligned}$$

To show the uniqueness of Nash equilibrium, it suffices to show that the linear system (7.39) admits a unique solution. We observe that the matrix A can be written as: $A = \Gamma_1(I - T_1)$ with $\Gamma_1 = diag(2b_1 - \delta_{\{\lambda^i - \alpha^i = 0\}}\frac{b_1^2}{2v_1}, ..., 2b_N - \delta_{\{\lambda^i - \mu^i = 0\}}\frac{b_N^2}{2v_N})$, and

$$T_1 = \begin{pmatrix} 0 & \vdots & \frac{c_{1N} - \delta_{\{\lambda^i - \mu^i = 0\}}\gamma_{1N}\frac{\beta_N}{2v_N}}{2b_1 - \delta_{\{\lambda^i - \mu^i = 0\}}\frac{\beta_1^2}{2v_1}} \\ & \ddots & \\ \frac{c_{N1} - \delta_{\{\lambda^i - \mu^i = 0\}}\gamma_{N1}\frac{\beta_1}{2v_1}}{2b_N - \delta_{\{\lambda^i - \mu^i = 0\}}\frac{\beta_N^2}{2v_N}} & \vdots & 0 \end{pmatrix}$$

We observe that the matrix T is substochastic since

$$\frac{\sum_{j\neq i} c_{ij} - \delta_{\{\lambda^i - \mu^i = 0\}}\gamma_{ij}\frac{\beta_j}{2v_j}}{2b_i - \delta_{\{\lambda^i - \mu^i = 0\}}\frac{\beta_i^2}{2v_i}} \leq \frac{\sum_{j\neq i} c_{ij}}{2b_i - \frac{\beta_i^2}{2v_i}} \leq \frac{b_i}{2b_i - \frac{\beta_i^2}{2v_i}} < \frac{b_i}{b_i} = 1.$$

The third inequality follows from Assumption 5. Thus $A^{-1} = (I - T_1)^{-1}\Gamma_1^{-1}$ and

$$\begin{aligned} p_i^* &= \sum_{j\in\mathcal{I}}(A^{-1})_{ij}(b_j v_j + a_j + \delta_{\{\lambda^j-\alpha^j=0\}}\frac{\beta_j^2}{2} - \sum_{k\neq j}\delta_{\{\lambda^k-\alpha^k=0\}}\gamma_{jk}\frac{\beta_k}{2} \\ &\quad +\beta_j(\delta_{\{\lambda^j>0\}}f_j^{\max} + \delta_{\{\alpha^j>0\}}f_j^{\min}) \\ &\quad -\sum_{k\neq i}\gamma_{jk}(\delta_{\{\lambda^j>0\}}f_k^{\max} + \delta_{\{\alpha^k>0\}}f_j^{\min})). \end{aligned}$$

We conclude that $(\mathbf{p}^*, \mathbf{f}^*)$ is the unique solution of Kuhn-Tucker conditions (7.33)-(7.34) and therefore the unique Nash equilibrium to the price-QoS game (7.9). □

7.9 Case study II: Loss or rejection probability as QoS

In this section, the quality of service corresponds to the loss probability, where one may model packet loss probabilities (e.g. at some common input queue) or rejection probability in a loss network (such as a circuit switched network).

$$f_i = (1 - P_{loss}^i)^{1/s}, \tag{7.40}$$

where P_{loss}^i is the loss probability. i.e,.

$$P_{loss}^i := G(\rho_i)$$

with, as usual, $\rho_i = d_i/\mu_i$ is the traffic intensity, and $s \geq 1$ is a scaling coefficient that adjusts the relative importance of the QoS parameter with respect to price. Note that when $s = 1$, QoS enters linearly in the demand function, D, whereas for $s > 1$, the QoS increases as a concave function of its parameters; in other words, for $s > 1$, we allow for *decreasing rates of return* on the quality of service provided.[3]

We will use the notation $G'(\rho)$ to denote the derivative with respect to ρ and we assume that $G'(\rho)$ is positive function over the interval $(0, +\infty)$. Hence, the loss probability is strictly increasing in ρ, and the function G^{-1} exists. Consequently, for all $i \in \mathcal{I}$,

$$g_i(f_i) = \frac{1}{G^{-1}(1 - f_i^s)} \tag{7.41}$$

Indeed, since G^{-1} exists, then $d_i/\mu_i = \rho_i = G^{-1}(1 - f_i)$, it follows that $\mu_i = \frac{1}{G^{-1}(1-f_i)}d_i$. Hence from (7.14), the function g_i is given by (7.41). Note that $f_i^{\min} = 0$ and $f_i^{\max} = 1$ for every $i \in \mathcal{I}$.

We are going to present some important special cases for which the assumption A4 is verified. In 7.9.1 and 7.9.3 we consider only the linear relationship between QoS and loss probability, that is $s = 1$, whereas in 7.9.2 we give results for strictly concave f as well.

7.9.1 Packet loss probability: a basic M/M/1/K queuing model

In this paragraph, the quality of service corresponds to the loss probability, or probability of a full buffer, in the M/M/1/K queue. We note that this could be a good approximation for the loss probabilities if the service provider offers access to a network, and the bottleneck in terms of losses is in some common input buffer to that network. Thus,

$$G(\rho_i) := \begin{cases} (1-\rho_i)\rho_i^{K_i}/(1-\rho_i^{K_i+1}) & \text{if} \quad 0 \leq \rho_i < 1,\ K_i = 1,2,3,.., \\ 1/(K_i+1) & \text{if} \quad \rho_i = 1,\ K_i = 1,2,3,.., \end{cases} \tag{7.42}$$

[3] *Decreasing rates of return* means that the marginal benefit of QoS decreases with increasing QoS levels This phenomenon is observed often in practice; the amount that a user is willing to pay, per unit of QoS, decreases as the level increases.

where K_i is the buffer size. Clearly, the derivative $G'(\rho)$ is positive, hence we can use the above framework. Let the concavity coefficient satisfy $s = 1$, implying that QoS measure f and loss probabilities are linearly related [see (7.40)].

Example 3. *For $K_i = 2$, g_i is*

$$g_i(f_i) = \frac{2f_i}{1 - f_i + \sqrt{1 + 2f_i - 3f_i^2}}$$

The next lemma characterizes some properties of the function g_i with respect to service levels f_i, $i \in \mathcal{I}$.

Lemma 2. *The function $g_i(.) \in C^2$ is convex, increasing, and its derivative satisfies*

$$\lim_{f_i \to 1} g_i'(f_i) = +\infty, \text{ and } \lim_{f_i \to 0} g_i'(f_i) = 1.$$

Proof: See Appendix. □

The above result implies that the function g_i verifies Assumption 4. Hence, from Theorem 6, we have the following result.

Corollary 1. *The price-QoS game (7.9) has a unique two-parameter Nash equilibrium $(\mathbf{p}^*, \mathbf{f}^*)$, with f_i^* the unique solution to the equations*

$$\begin{aligned} g_i'(f_i^*) &= \frac{\beta_i}{v_i b_i} \text{ if } \frac{\beta_i}{v_i b_i} \geq 1 \\ f_i^* &= 0 \text{ if } \frac{\beta_i}{v_i b_i} < 1 \end{aligned}$$

and $\mathbf{p}^*$ is then obtained through (7.15).

7.9.2 Packet loss probability: incorporating large-deviation scaling

Let P_{loss} be, as above, the loss probability for a M/M/1/K queue with K the buffer size at the queue. Given the probability of loss from (7.42), where $\rho_i := \frac{d_i}{\mu_i}$, and letting the arrival rate, capacity and buffer size be scaled by a factor n, as in many-source large-deviation scaling, we obtain, by letting $n \to \infty$,

$$\lim_{n \to \infty} \frac{(1 - \rho_i)\rho_i^{nK}}{1 - \rho_i^{nK+1}} = \frac{(d_i - \mu_i)^+}{d_i}, \tag{7.43}$$

with $(\cdot)^+ = \max\{0, \cdot\}$. Note that even though the buffer size goes to ∞, the delay remains constant as the capacity, μ_i, also goes to ∞. In a deterministic fluid model, this has the simple interpretation of fraction of fluid lost when the arrival rate exceeds capacity.

Using (7.42), the quality of service is then given by

$$f_i = \left[1 - \frac{(d_i - \mu_i)^+}{d_i}\right]^{1/s}. \tag{7.44}$$

Note that (7.44) is not a differentiable function due to the max operation. However, since the QoS measure remains constant ($f_i = 1$) when the capacity exceeds the demand, it is sufficient to consider that the largest capacity implemented would be $\mu_i = d_i$. Hence, we shall neglect the max operation in (7.44), and then the bandwidth required in order to guarantee the QoS f_i is given by:

$$\mu_i = g_i(f_i)d_i \text{ with } g_i(f_i) = f_i^s$$

The profit of **SP** i is in turn given by:

$$U_i(\mathbf{p},\mathbf{f}) = D_i(\mathbf{p},\mathbf{f})(p_i - v_i f_i^s) \tag{7.45}$$

where the strategy space is given by the subset $\mathcal{R}_i$ where

$$\mathcal{R}_i = \{(p_i, f_i) : \ p_i^{\min} \le p_i \le p_i^{\max};\ 0 \le f_i \le 1\}$$

We shall require the following assumption, which is likely to hold, since none of the constants β, v or b should be zero in this setting.

Assumption 6. *The constants b, v, and β satisfy:*

$$\beta_i - b_i v_i \neq 0 \tag{7.46}$$

Theorem 8. *The price-QoS problem (7.9) with Quality of Service $\mathbf{f}$ as in (7.44) has a unique two parameter Nash equilibrium $(\mathbf{p}^*, \mathbf{f}^*)$, with f_i^* is given by:*

1. *For concave QoS coefficient, $s > 1$, that is, under decreasing rate of return on the QoS,*

$$f_i^* = \begin{cases} \left(\frac{\beta_i}{s v_i b_i}\right)^{\frac{1}{s-1}} & \text{if} \quad \frac{\beta_i}{s v_i b_i} < 1, \\ 0 & \text{if} \quad \frac{\beta_i}{s v_i b_i} \ge 1. \end{cases}$$

2. *For $s = 1$, that is, under constant rate of return on the QoS level, $\beta_i \neq b_i v_i$, $i \in \mathcal{I}$, so that*

$$f_i^* = \left\{ \begin{array}{lll} 1 & \text{if} & \beta_i < v_i b_i, \\ 0 & \text{if} & \beta_i > v_i b_i. \end{array} \right\}$$

Proof: For all $s > 0$, the function g_i is convex, this implies that the function U_i is concave. Hence, we need only to show that that the profit function of each provider i is jointly log-concave in (p_i, f_i) for $s = 1$. Recall that $\tilde{U}_i$ denotes $\log U_i$. Note that

$$\begin{aligned}
\frac{\partial \tilde{U}_i(p,\mathbf{f})}{\partial p_i} &= \frac{1}{(p_i - v_i f_i)} - \frac{b_i}{D_i(p^*,\mathbf{f})} \text{ and } \frac{\partial \tilde{U}_i(p,\mathbf{f})}{\partial f_i} = -\frac{v_i}{(p_i - v_i f_i)} + \frac{\beta_i}{D_i(p^*,\mathbf{f})} \\
\frac{\partial^2 \tilde{U}_i(p,\mathbf{f})}{\partial p_i^2} &= \frac{-1}{(p_i - v_i f_i)^2} - \frac{b_i^2}{D_i^2(p^*,\mathbf{f})} < 0, \ \frac{\partial^2 \tilde{U}_i(p,\mathbf{f})}{\partial p_i \partial f_i} = \frac{v_i}{(p_i - v_i f_i)^2} + \frac{b_i \beta_i}{D_i^2(p^*,\mathbf{f})} \\
\frac{\partial^2 \tilde{U}_i(p,\mathbf{f})}{\partial f_i^2} &= \frac{-v_i^2}{(p_i - v_i f_i)^2} - \frac{\beta_i^2}{D_i^2(p^*,\mathbf{f})} < 0.
\end{aligned}$$

Since $\frac{\partial^2 \tilde{U}_i}{\partial f_i^2} < 0$ and $\frac{\partial^2 \tilde{U}_i}{\partial p_i^2} < 0$, it suffices to show that the determinant of the Hessian matrix H is positive, where

$$H = \begin{pmatrix} \frac{\partial^2 \tilde{U}_i}{\partial p_i^2} & \frac{\partial^2 \tilde{U}_i}{\partial p_i \partial f_i} \\ \\ \frac{\partial^2 \tilde{U}_i}{\partial f_i \partial p_i} & \frac{\partial^2 \tilde{U}_i}{\partial f_i^2} \end{pmatrix}$$

We have

$$\begin{aligned}
\det H[\tilde{U}_i(\mathbf{p},\mathbf{f})] &= \frac{\partial^2 \tilde{U}_i}{\partial f_i^2}\frac{\partial^2 \tilde{U}_i}{\partial p_i^2} - \left(\frac{\partial^2 \tilde{U}_i}{\partial p_i \partial f_i}\right)^2 = \Big(D_i(p_i - v_i f_i)\Big)^{-4} \\
&\quad \Big([D_i^2 + b_i^2(p_i - v_i f_i)^2][v_i^2 D_i^2 + \beta_i^2(p_i - v_i f_i)^2] - [v_i D_i^2 + b_i \beta_i (p_i - v_i f_i)^2]^2\Big) \\
&= \frac{(\beta_i - b_i v_i)^2}{D_i^2 (p_i - v_i f_i)^2} > 0
\end{aligned}$$

The latter inequality follows from Assumption 6.

Next, we shall establish the uniqueness of the equilibrium. Note first that any solution to the following Kuhn-Tucker conditions is a Nash equilibrium: For every $i \in \mathcal{I}$, there exist non-negative Lagrange multipliers α^i and λ^i such that

$$
\begin{aligned}
0 &= \frac{\partial \tilde{U}_i(p,\mathbf{f})}{\partial p_i} = \frac{1}{(p_i - v_i f_i^s)} - \frac{b_i}{D_i(p^*,\mathbf{f})} && (7.47)\\
\lambda^i - \alpha^i &= \frac{\partial \tilde{U}_i(p,\mathbf{f})}{\partial f_i} = -\frac{s v_i f_i^{s-1}}{(p_i - v_i f_i^s)} + \frac{\beta_i}{D_i(p^*,\mathbf{f})} && (7.48)\\
\alpha^i f_i &= 0, \;\; \lambda^i(1-f_i) = 0, && (7.49)
\end{aligned}
$$

where, as before, α^i and λ^i are the Lagrange multipliers corresponding to the min and max constraints on the values of f_i; note, also as before, that we do not need Lagrange multipliers for the constraints on p_i since p_i^* is obtained at a point where the partial derivative (7.33) is zero, possibly at the lower boundary, (7.16) and (7.19). By multiplying (7.47) by β_i and (7.48) by b_i, and summing the resulting equations, we obtain

$$b_i(\lambda^i - \alpha^i) = \frac{\beta_i - s v_i b_i f_i^{s-1}}{(p_i - v_i f_i^s)} \tag{7.50}$$

Now, we are going to detemine the QoS for $s > 1$. We have use the fact that the partial derivative $\frac{\partial \tilde{U}_i(p,\mathbf{f})}{\partial f_i}$ is positive. Hence we do not need the Langrage mulitiplier α^i for the constraint $f_i - f_i^{min} \geq 0$. According to this, (7.50) becomes

$$b_i \lambda^i = \frac{\beta_i - s v_i b_i f_i^{s-1}}{(p_i - v_i f_i^s)}$$

We deduce that

$$f_i^* = \begin{cases} \left(\frac{\beta_i}{s v_i b_i}\right)^{\frac{1}{s-1}} & \text{if} \quad \frac{\beta_i}{s v_i b_i} \leq 1, \\ 1 & \text{if} \quad \frac{\beta_i}{s v_i b_i} > 1. \end{cases}$$

Letting now $s = 1$. It follows, then, from (7.50) with $s = 0$ that

$$f_i^* = \begin{cases} 0 & \text{if} \quad \beta_i < v_i b_i \\ 1 & \text{if} \quad \beta_i \geq v_i b_i \end{cases}$$

Fixing $\mathbf{f} = \mathbf{f}^*$, we obtain that $\mathbf{p}^*$ is the *unique* solution of the linear system given by (7.21). Thus $(\mathbf{p}^*, \mathbf{f}^*)$ is consequently is the unique Nash equilibrium. □

7.9.3 Session rejection probability: the Erlang loss formula

Consider now a model handling blocked calls in which there are K servers available, and each newly arriving customer is given his own server; if a customer arrives when all servers are occupied, that customer's request is lost. This system is of particular interest in telephone networks, and is known as Erlang's loss formula. In terms of our notation, we have:

$$P_{loss}^i = G_i(\rho_i) = \frac{(\rho_i)^{K_i}/K_i!}{\sum_{j=0}^{K_i} (\rho_i)^j/j!} \tag{7.51}$$

We begin by taking the first derivatives of $G_i(.)$ with respect to ρ. After some simplification, we obtain

$$G_i'(\rho_i) = \frac{\sum_{j=0}^{K_i-1} (K_i - j)\rho_i^{K_i+j-1}/K_i!j!}{[\sum_{j=0}^{K_i} (\rho_i)^j/j!]^2} > 0 \tag{7.52}$$

Hence the function G_i^{-1} exists.

Let the concavity coefficient, $s = 1$ implying that QoS measure f and loss probabilities are linearly related (see (7.40)). The following lemma characterizes some properties of the function g_i with respect to f_i.

Lemma 3. *The function $g_i(.) \in C^2$ is convex, increasing, and its derivative satisfies*

$$\lim_{f_i \to 1} g_i'(f_i) = +\infty, \text{ and } \lim_{f_i \to 0} g_i'(f_i) = \frac{1}{K_i}$$

Proof: See Appendix. □

Combining the last lemma with Theorem 6 we obtain the following result.

Corollary 2. *The two-parameter price-QoS game (7.9), with QoS defined as in (7.42) with $s = 1$ and session loss probability given by (7.51), has a unique Nash equilibrium $(\mathbf{p}^*, \mathbf{f}^*)$, with f_i^* the unique solution to the equations*

$$\begin{aligned} g_i'(f_i^*) &= \frac{\beta_i}{v_i b_i} \text{ if } K_i \geq \frac{v_i b_i}{\beta_i} \\ f_i^* &= 0 \text{ if } K_i < \frac{v_i b_i}{\beta_i} \end{aligned}$$

and $\mathbf{p}^*$ is obtained through (7.15).

7.10 Appendix

Proof of Lemma 2

We require the second derivative of $g_i(.)$ with respect to f_i. Hence we have

$$g_i'(f_i) = \frac{[G_i^{-1}(1-f_i)]'}{\{G_i^{-1}(1-f_i)\}^2}, \tag{7.53}$$

and

$$g_i''(f_i) = \frac{-[G_i^{-1}(1-f_i)]''\{G_i^{-1}(1-f_i)\}^2 + 2\{[G_i^{-1}(1-f_i)]'\}^2[G_i^{-1}(1-f_i)]}{\{G_i^{-1}(1-f_i)\}^4} \tag{7.54}$$

$$[G_i^{-1}(x)]' = \frac{1}{G_i'(G_i^{-1}(x))}, \text{ and } [G_i^{-1}(x)]'' = -\frac{G_i''(G_i^{-1}(x))}{\{G_i'(G_i^{-1}(x))\}^3} \tag{7.55}$$

We will find that the second derivative of $g_i(.)$ is positive. Since $G_i(.)$ is increasing function, then G_i^{-1} is also increasing, thus from (7.53), $g_i^{-1}(.)$ is also increasing function. On the other hand, we have Theorem 3 in [?] states that the loss probability $G_i(\rho)$ is a convex of μ_i with d_i fixed, then from (7.55), the concavity of G_i^{-1} is immediate. Thus from (7.54), the convexity of g_i is immediate.

Let us now calculate the limit of g_i' as $f_i \to 1$. Working with equations (7.53) and (7.55), we have

$$\Phi(a) := \lim_{f_i \to a} g'(f_i) = \lim_{f_i \to a} \left[\frac{1}{\{G_i^{-1}(1-f_i)\}^2} \frac{1}{G_i'(G_i^{-1}(1-f_i))} \right] \tag{7.56}$$

Making the change of variable $G_i^{-1}(1-f_i) = X$ we have

$$\Phi(a) := \lim_{X \to G_i^{-1}(1-a)} \frac{1}{X^2 G_i'(X)} = \lim_{X \to G_i^{-1}(1-a)} \frac{(1-X^{K+1})^2}{X^{2K+2} - (K+1)X^{K+2} + KX^{K+1}}$$

Further, it is clear that since $G_i(0) = 0$ then $G_i^{-1}(0) = 0$. Combining the last two we have

$$\Phi(1) = \lim_{f_i \to 1} g'(f_i) = \lim_{X \to 0} \frac{X^{2K+2} - 2X^{K+1} + 1}{X^{2K+2} - (K+1)X^{K+2} + KX^{K+1}} = +\infty$$

Let us now show the limit of g'_i as $f_i \to 0$. We see that $\lim G_i^{-1}(y) = +\infty$ as $y \to 1$ since the limit of $G_i(\rho)$ as $\rho \to +\infty$ is 1. Consequently,

$$\Phi(0) = \lim_{f_i \to 0} g'(f_i) = \lim_{X \to +\infty} \frac{X^{2K+2} - 2X^{K+1} + 1}{X^{2K+2} - (K+1)X^{K+2} + KX^{K+1}} = \lim_{X \to +\infty} \frac{X^{2K+2}}{X^{2K+1}} = 1$$

which proves the desired result. □

Proof of Lemma 3

The Proposition 3 in [?] states that the Erlang loss formula is strictly convex in μ_i, keeping d_i fixed. Hence the concavity of g_i follows along similar lines.

Let us now calculate the limit of g'_i as $f_i \to 1$. Similarly, we have

$$\Phi(1) = \lim_{f_i \to 1} g'(f_i) = \lim_{X \to 0} \frac{(\sum_{j=0}^{K_i} X^j/j!)^2}{X^2(\sum_{j=0}^{K_i-1}(K_i - j)X^{K_i+j-1}/(K_i!j!))} = +\infty$$

and

$$\Phi(0) = \lim_{f_i \to 0} g'(f_i) = \lim_{X \to +\infty} \frac{(\sum_{j=0}^{K_i} X^j/j!)^2}{X^2(\sum_{j=0}^{K_i-1}(K_i - j)X^{K_i+j-1}/(K_i!j!))} \quad (7.57)$$

$$= \lim_{X \to +\infty} \frac{X^{2K_i}/(K_i!)^2}{X^{2K_i}/(K_i!(K_i-1)!)} = \frac{1}{K_i} \quad (7.58)$$

which proves the desired result. □

7.11 Conclusions

We have presented a framework for modeling the complex interactions among telecom and Internet service providers through a class of one- and two-parameter Nash equilibrium models. The novel properties of the models are the use of demand functions which describe each service providers' customer set, and take into account not only the characteristics of that service provider, (**SP**), i, but also of all other **SP**, j, the presence of two parameters describing each **SP**'s service price and Quality of Service (QoS) level, and the generality of the utility functions, which permit incorporating complex functions such as those describing loss probabilities, decreasing rate of return, and other general QoS forms.

The following potentially interesting avenues for further research have arisen during this study.

- It could be valuable to consider nonlinear expressions for the demand experienced by each **SP**. One could envisage some general concave functions, or the logit or probit models, for example. The logit model has the positive feature of allowing closed form expressions for the choice probabilities across **SP**, but suffers the drawback of requiring an i.i.d. assumption on those probabilities. The probit model allows for covariances across **SP**s, but requires numerical simulation, in general, to obtain solutions.

- It could be of interest to incorporate explicitly the network of each **SP** into the model. Indeed, in this work, the characteristics of each provider's network have been aggregated into a single pair of functions, representing some sort of *average* service one would experience using that service provider. However, it is well known that QoS, and in some cases, price, vary considerably across the paths of a single **SP**. In addition, the demand experienced by an **SP** depends naturally on the source and destination of the demand.

Chapter 8

Potential games and Equivalent Games

8.1 Equivalent Games: How to Transform a Game

In this section we identify some transformations of a game that keep its equilibrium strategies unchanged.

Theorem 9. *Consider an I-player game. $\mathcal{A}^i$ is the set of actions of player i and $U^i(a)$ is his utility function, where a is a multistrategy. Consider for any i one of the following transformations:*

- *Adding functions of the actions of other players:*

$$\widehat{U}^i(a) := U^i(a) + f^i(a_{(-i)})$$

  where f^i is an arbitrary function of the actions of all players other than i.

- *Composing with a monotone increasing function:*

-

$$\widehat{U}^i(a) := g(U^i(a))$$

a^* is an equilibrium for the game with utilities U^i if and only if is an equilibrium for game with utilities $\widehat{U}^i$.

8.1.1 Example: transforming a game into a global optimization problem

As an example, consider a two player multi-access game where each player i has two actions: transmit denoted by "1" or listen denoted by "0". If both transmit then there is a collision and the packet is lost. Assume that both players wish to maximize the global throghput, given by

$$\Theta = a_1 a_2$$

where a_1 and a_2 are the actions of the two players. In addition, each player has a cost of h per energy. The utilities for the players are thus given by

$$U^i(a_1 a_2) = a_1 a_2 - h a_i$$

According to the above Theorem, it follows that the equilibria for this game conincide with the one in which we add to each player $i = 1, 2$ the extra utility $-h a_j$ for $j \neq i$. After this modification, the utilities of the players become the same, both wishing to maximize

$$U(a_1 a_2) = a_1 a_2 - h(a_1 + a_2)$$

We thus transformed the game into a global optimization problem. A solution to the global optimization problem is obviously a Nash equilibrium for that problem. By the Theorem above it is thus an equilibrium for the original game.

8.1.2 Example: transforming a non zero-sum game into a zero-sum game

Consider the same Example as in the previous section but this time with

$$U^1(a_1a_2) = a_1a_2 - ha_1,$$

$$U^2(a_1a_2) = -a_1a_2 - ha_2,$$

In other words, player 2 tries to cause harm and to minimize the throughput that the other player tries to maximize. Both players have in addition, energy costs.

Consider next the transformed utilities:

$$\widehat{U}^1(a_1a_2) = a_1a_2 - ha_1 + ha_2,$$

$$\widehat{U}^2(a_1a_2) = -a_1a_2 - ha_2 + ha_1,$$

Now the utilities sum to zero. We thus transformed the game into a global zero-sum game. A saddle point for the zero-sum game thus provides us an equilibrium for the original game.

8.2 Exercises

1. Compute the equilibrium and the utility at equilibrium in the Multiple-Access with Capture when pricing the high power.

2. Consider a two person game with performance measures U^1 and U^2. Define for $\alpha \in [0,1]$ the corresponding cooperative utility with cooperation degree α:

$$V^1(a) = \alpha U^1(a) + (1-\alpha)U^2(a).$$

Note that for $\alpha = 1$ we get the standard non-cooperative game, for $\alpha = 0.5$ we get the team problem, for $\alpha = 0$ we have the "altruistic" problem, where player 1 wishes to maximize the performance of player 2. Analyze the Multiple-Access game when the utility of player 1 is $V^1(a)$ and where

- The utility of Player 2 is $V^2(a)$,
- The utility of Player 2 is $U^2(a)$,

Chapter 9

Measuring the Quality of an Equilibria

In this lecture, we focus on measuring and defining efficiency of points in the utility set of a given game. Consider for example the following game in pure strategies:

	Choice A	Choice B
Choice A	(00, 10)	(20, 79)
Choice B	(79, 20)	(50, 50)

If you were a system operator and could dictate or advise the two users, what would you tell them? Intuitively, the set of choices (B,B) seems like an interesting one, as it is *fair* in the sense that the two users would receive the same utility. On the other hand, the choices (A,B) and (B,A) are also of interest since they would give one of the two players a higher payoff and a comparable global payoff.

Consider a normal form game $(\mathcal{N}, \mathcal{S}, \mathcal{U})$. The set of players is $\mathcal{N}$ with their strategy space $\mathcal{S} = \prod_{n \in \mathcal{N}} \mathcal{S}_n$ and $\mathcal{S}_n$ be the possible strategies of player n (note that this set can be discrete or continuous). Finally, $\mathcal{U} = \prod_{n \in \mathcal{N}} \mathcal{U}_n$ is the utility set functions, that is $\mathcal{U}_n$ is a set of mappings from $\mathcal{S}$ to a totally ordered space (typically the set of real numbers $\mathbb{R}$). Yet, even though there exists a total order on the utilities sets of each player, there is no natural and compatible total order on their product, making comparisons between utility points hazardous.

Appropriate ways to define an optimal point in the context of systems shared by several users is the topic of this lecture. It is strictly analogous to simultaneously optimizing several objectives in a single user system. This second approach has been extensively studied in the context of *multi-criteria optimization.* In this lecture, we start by reviewing a few techniques used in multi-criteria optimization and their equivalent in game theory (Section **??**). We then present the definition of Pareto optimality, based on the canonical partial order on $\mathbb{R}^n$ and its properties (Section **??**). We also discuss its connections with the classical definitions in multi-criteria optimization. Finally, we discuss possible measures of the inefficiency, based on the distance of a given point (Nash equilibrium) to the Pareto border (Section 9.5).

9.1 Pareto Efficiency

The notion of Pareto optimality is based on the extension of the canonical order to $\mathbb{R}^n$.

9.1.1 Definition

Definition 7 (Canonical partial orders)**.** *We consider the following orders as being canonical.*

- *The canonical partial $\preccurlyeq$ order on $\mathbb{R}_+^n$ is defined by:*

$$u \preccurlyeq v \Leftrightarrow \forall k : u_k \leq v_k$$

The classical strict partial order $\prec$ is defined accordingly. ($u \preccurlyeq v \Leftrightarrow u \prec v$ and $u \neq v$.)

- *We also define an additional strict partial order $\ll$ on $\mathbb{R}_+^n$, namely the strict Pareto-superiority, by:*

$$u \ll v \Leftrightarrow \forall k : u_k < v_k$$

We can then define the Pareto set as the optimal points for $\preccurlyeq$:

Definition 8 (Pareto optimality). *A choice $u \in U$ is said to be Pareto optimal if*

$$\forall v \in U, \exists i, v_i > u_i \Rightarrow \exists j, v_j < u_j.$$

In other words, u is Pareto optimal if it is maximal in U for the canonical partial order on $\mathbb{R}_+^n$.

Roughly speaking, a point is said to be Pareto optimal if it is impossible to increase the performance of a criteria without decreasing that of another one. With the previous notations, this can be written: x is Pareto optimal iff $\forall y, \exists i \in 1..I, u_i(y) > u_i(x) \Rightarrow \exists j \in 1..I, u_j(y) < u_j(x)$. In other words, a Pareto optimal point is such that there exists no point that is greater, component-wise.

A policy function is said to be Pareto-optimal *if for all $U \in \mathcal{U}, \alpha(U)$ is Pareto-optimal.*

The key idea here is that Pareto optimality is a global notion. Even in systems that consists of independent elements, the Pareto optimality cannot be determined on each independent subsystem. Such phenomena has been exhibited in [47]. The considered system is a master-slave platform in which the master can communicate with as many slaves as it needs at any time. The master holds a infinite number of tasks corresponding to N applications, and each of them can be executed on any slave. The authors study the system at the Nash equilibrium (each application competing with each other for both resource and CPU). Although the problems associated with each machine is independent, the authors show that for any system with one slave the equilibrium is Pareto optimal, while Pareto inefficiency can occur in multiple slave systems (see Figure 9.1 and 9.1.1).

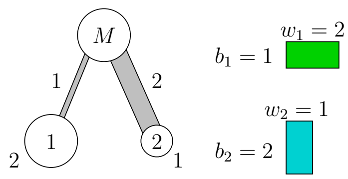


Figure 9.1: A master-slave system: 2 applications and 2 computing slaves

9.1.2 Properties

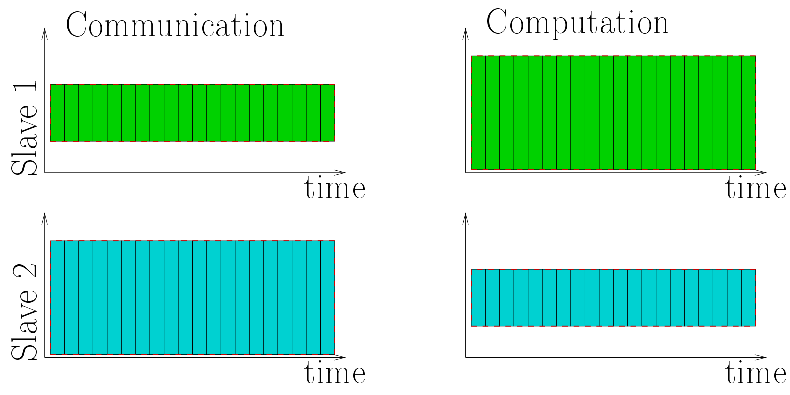
Pareto Optimality and Aggregation Function

Definition 9 (f-increasing). *A policy α is said to be f-*increasing *if $f \circ \alpha$ is monotone. Any f-optimizing policy is thus f-increasing.*

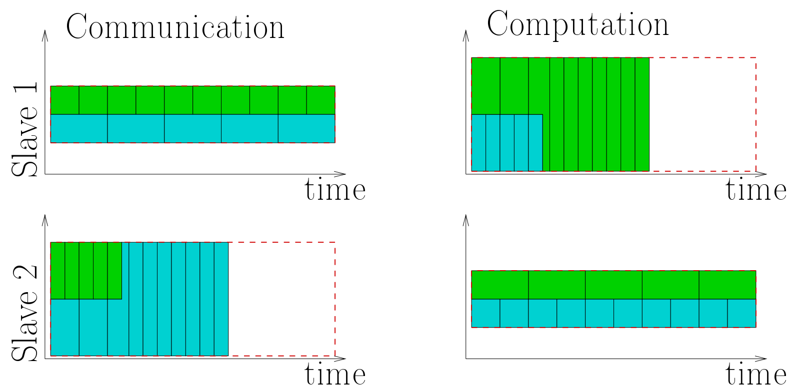
Pareto optimality and monotony of the index optimization are closely related, as illustrated in the following results.

Theorem 10. *Let α be an f-optimizing policy. If f is strictly monotone then α is Pareto-optimal.*

Proof. Suppose that α is not Pareto optimal. Then, there exists $U = \mathcal{U}(\mathcal{S})$ such that $\alpha(U)$ is not Pareto optimal. Hence, there exists $v \in U$ such that $\alpha(U) \prec v$, and hence $f(\alpha(U)) < f(v)$, which contradicts the definition of $\alpha(U)$. $\square$



(a) A case of a single master and two slaves system: the throughput achieved by this allocation is 1 for the two applications (green and blue)



(b) Same machines and applications: with this allocation (which is the Nash Equilibrium), each application gets a throughput of 3/4. If green is numbered 1 and blue is 2, the allocation is $(1/2, 1/4)$ on the first slave and $(1/4, 1/2)$ on the second slave. Hence, if each slave was alone, the allocation would be Pareto optimal.

Theorem 11. *Let α be an f-optimizing policy. If α is Pareto-optimal then f is monotone.*

Proof. Suppose that f is not monotone. Then there exists $u \prec v$ such that $f(v) \prec f(u)$. Consider $U = \{u, v\}$. As $u \prec v$ and α is Pareto-optimal, then $\alpha(U) = v$ which is in contradiction with $f(v) \prec f(u)$. $\square$

Arithmetic and geometric mean indexes are examples or strictly monotone index. On the other hand, the Jain index is an example of non-monotone index. The min and the max index are also not strictly monotone, which is why, max-min fairness or min-max fairness are recursively defined in the literature (see lecture on bargaining).

Continuity

Let $\mathcal{H}(\mathbb{R}_+^n)$ denote the set of non-empty compact sets of $\mathbb{R}_+^n$ and $\mathcal{C}$ denote the set of non-empty compact and convex sets of $\mathbb{R}_+^n$. The canonical partial order on $\mathcal{H}(\mathbb{R}_+^n)$ is the classical inclusion order: $\subseteq$.

In the following, we assume that $\mathcal{U}(\mathcal{S})$ the set of all utility sets is either equal to $\mathcal{H}(\mathbb{R}_+^n)$ or $\mathcal{C}(\mathbb{R}_+^n)$. Any negative result regarding $\mathcal{C}(\mathbb{R}_+^n)$ also applies to $\mathcal{H}(\mathbb{R}_+^n)$.

To study the continuity of policy functions, we need a topology on $\mathcal{U}(\mathcal{S})$. That is why in the following, we use the classical metric on compact sets.

Definition 10 (Hausdorff metric)**.** *Considering a metric function d on $\mathbb{R}_+^n$, one can define the* distance from x to *the compact* B *as:*

$$d(x,B) = \min\{d(x,y)|y \in B\}$$

The distance from *the compact* A to *the compact* B *as:*

$$d(A,B) = \max\{d(x,B)|x \in A\}$$

The Hausdorff distance *between two compacts* A *and* B *can thus be defined as:*

$$h(A,B) = \max(d(A,B), d(B,A))$$

$(\mathcal{H}(\mathbb{R}^n_+), h)$ and $(\mathcal{C}(\mathbb{R}^n_+), h)$ are complete metric spaces [48] and we can thus study the continuity of policy functions under pretty clean conditions.

Let us assume that a configuration r made of p resources is modeled as an element r of $\mathbb{R}^p_+$. A set of configurations R is thus a compact of $\mathcal{H}(\mathbb{R}^p_+)$. Let us assume that utility of users g are continuous functions from $\mathbb{R}^p_+$ to $\mathbb{R}^n_+$. Then utility sets $U = \mathcal{U}(\mathcal{S})$ are built with the help of R and g.

$$U : \begin{cases} (\mathcal{H}(\mathbb{R}^p_+), C_0(\mathbb{R}^p_+, \mathbb{R}^n_+)) & \to \mathcal{H}(\mathbb{R}^n_+) \\ (R, g) & \mapsto \{g(r)|r \in R\} \end{cases}$$

The mapping U being continuous, $\alpha \circ U$ represents the sensibility of the allocation with respect to resources and utility functions. Continuity of the allocation α is thus an essential feature. Indeed, it ensures that a slight change in the system resources would not significantly affect the allocation. In dynamically changing systems, this ensures a certain stability. It also ensures that a slight error in utility functions does not affect too much the allocation.

Theorem 12. *The Pareto set of a convex utility set is not necessarily compact.*

The function $\overline{\mathcal{P}}$ *from* $\mathcal{C}(\mathbb{R}^n_+)$ *to* $\mathcal{H}(\mathbb{R}^n_+)$ *that associates to* U *the closure of its Pareto set is not continuous.*

Proof. Let us first exhibit a convex utility set whose Pareto set is not closed. Let $C = \{(x,y,z) \in \mathbb{R}^3_+ | x^2 + y^2 \leq 1, 0 \leq z \leq 1 - \frac{1}{2}\frac{x^2 - 2y^2 + 1}{1-y}\}$. The set C is depicted on Figure 9.2. We have $\overline{\mathcal{P}}(C) = \{(x, y, 1 - \frac{1}{2}\frac{x^2-2y^2+1}{1-y}) | x^2 + y^2 \leq 1, x + y \geq 1, x > 0\} \cup \{(0,1,1)\}$, which is not closed.

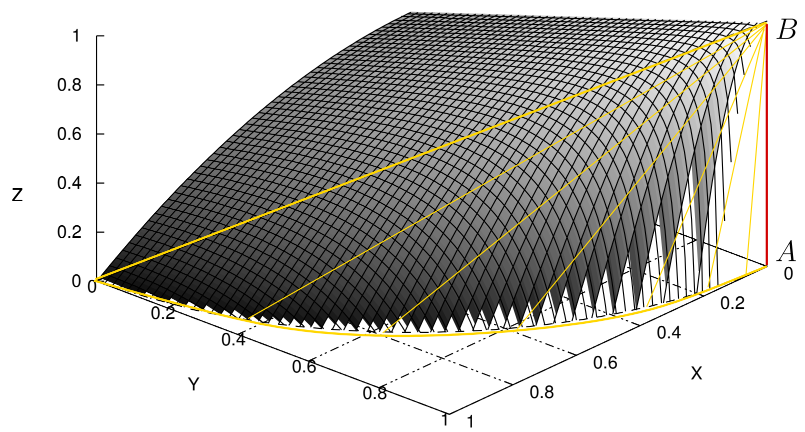


Figure 9.2: Convex set whose Pareto set is not closed. The segment $[A, B[$ does not belong to the Pareto set.

Let us consider $\overline{\mathcal{P}}$ from $\mathcal{C}(\mathbb{R}^n_+)$ to $\mathcal{H}(\mathbb{R}^n_+)$ that associate to U the closure of its Pareto set. Figure 9.3 depicts a converging sequence of convex C_n such that $\overline{\mathcal{P}}(C_n)$ does not converge to $\overline{\mathcal{P}}(C_\infty)$. □

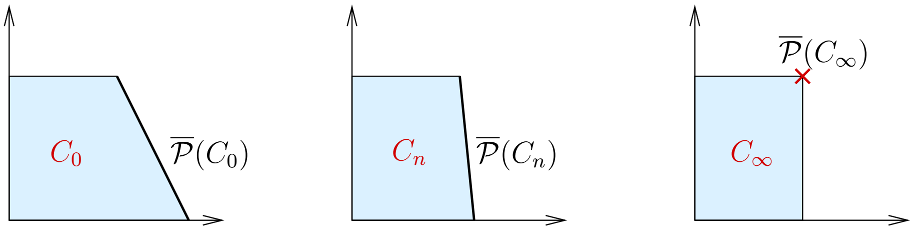


Figure 9.3: $\overline{\mathcal{P}}$ is not continuous.

Theorem 13. *Let α be a general Pareto-optimal policy function. α is not continuous.*

Proof. We prove that α cannot be continuous with the simple instances depicted on Figure 9.4. The only Pareto-optimal points are A and B. Therefore α has to choose in the first set between A and B. If A_1 is chosen, then by moving A_1 to A_0, the choice has to "jump" to B, hence α is not continuous. □

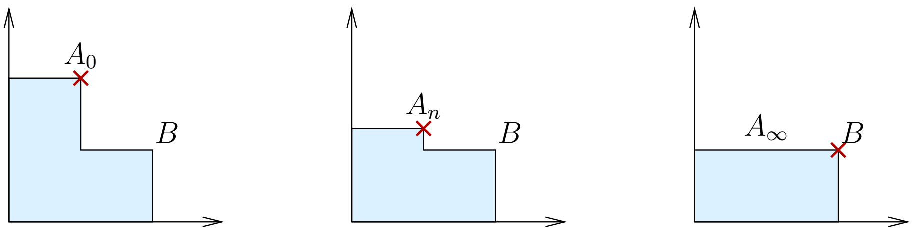


Figure 9.4: General Pareto-optimal policies are discontinuous: a path leading to discontinuity.

Remark 7. *There exists continuous and non-continuous convex Pareto-optimal policy functions.*

Proof. Let us consider a policy function α optimizing the sum of utilities. The two convex sets on Figure 9.5 show that α is not continuous around the set $K = \{(x, y)|x+y \leq 1\}$. This discontinuity is due to the fact that many different points of K simultaneously optimize the sum.

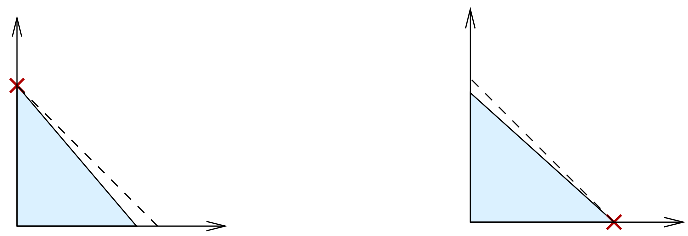


Figure 9.5: Optimizing $\sum$: a discontinuous convex policy.

The policy function α optimizing the product $\prod$ of utilities is continuous though. As $\prod$ is strictly monotone, α is Pareto-optimal. Moreover, as for any c, $I_c = \{x \in \mathbb{R}_+^n | \prod x_i \geq c\}$ is strictly convex, for any convex, there is a single point optimizing the $\prod$. Let us assume by contradiction that α is not continuous at the point C. Then there exists C_n converging to C and such that $x_n = \alpha(C_n)$ converges to $x_\infty \neq \alpha(C)$. As our sets are compact, there exists a sequence $y_n \in C_n$ such that y_n converges to $\alpha(C)$. By definition, we have $\forall n, \prod(y_n) \leq \prod(x_n)$. Therefore $\prod(\alpha(C)) \leq \prod(x_\infty)$, which is absurd as $\alpha(C)$ is optimal in C for $\prod$ and $\alpha(C) \neq x_\infty$. □

Monotony

We state in this sub-section two results on monotonicity of index and policy functions. The first one emphasizes that index-functions only measure a specific characteristic of performance measure, and are hence not compatible. This explains why allocations that are efficient (optimizing the arithmetic mean) cannot (in general) also be fair (optimizing the geometric mean).

The second result states that, even when restricted to convex utility sets, policy functions cannot be monotone. This infers that even in Braess-free systems (see lecture on routing games), an increase in the resource can be detrimental to some users.

Definition 11 (Braess-paradox). *A policy function α is said to have Braess-paradoxes it there exists U_1 and U_2 such that*

$$U_1 \subset U_2 \text{ and } \alpha(U_1) \gg \alpha(U_2)$$

with $\gg$ defined as in definition 7. A policy function such that there is no Braess-paradox is called Braess-paradox-free.

Theorem 14. *Let f and g be two monotone index functions. A g-optimizing policy α_g is f-increasing if and only if α_g is f-optimizing.*

Proof. If α_g is f-optimizing, then α_g is clearly f-increasing.

Let us assume that α_g is not f-optimizing. We define the partial order $\prec_f$ (resp. $\prec_g$) on $\mathbb{R}_+^n$ by $x \prec_f y$ iff $f(x) \leq f(y)$. We have $\prec_f \neq \prec_g$, otherwise α_g would be f-optimizing. Thus there exists $\overline{x_1}$ and $\overline{x_2}$ such that: $\overline{x_1} \prec_f \overline{x_2}$ and $\overline{x_2} \prec_g \overline{x_1}$. Considering $U = \{x_1\}$ and $U' = \{x_1, x_2\}$, shows that α_g is not f-increasing. $\square$

In other words, a policy optimizing an index f is always non-monotone for a *distinct* index g.

Theorem 15. *Even if convex, policy functions cannot be monotone.*

Proof. Let us consider α a monotone convex policy function and let us consider the three following convex sets $U_1 = \{(0,1)\}$, $U_2 = \{(1,0)\}$, and $U_3 = \{(x, 1-x) | 0 \leq x \leq 1\}$ (see Figure 9.6).

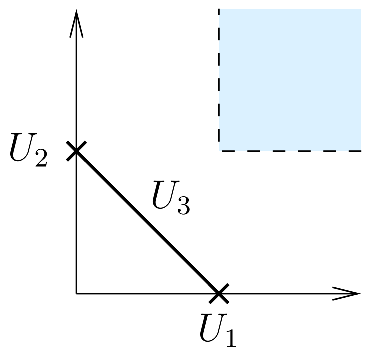


Figure 9.6: Even convex policy functions cannot be monotone.

We necessarily have $\alpha(U_1) = (0,1)$ and $\alpha(U_2) = (1,0)$. As U_1 and U_2 are subsets of U_3, we have $\alpha(U_3) \succcurlyeq (1,1)$, which is absurd because no such point belongs to U_3. $\square$

9.1.3 Conclusion

In this section, we have established the following results:

- Indexes should be strictly monotone to ensure Pareto-Optimality.
- Continuity (of allocations) is only possible when considering convex utility sets.

- It is impossible to ensure that the growth of the utility set does not incur the decrease of the utility of some player (i.e. policy functions cannot be monotone, even when restricting to convex utility sets).
- A policy optimizing a given index f leads to erratic values of an other index g when growing utility sets (unless f and g induce the same optimization).

Note that even though being Braess-paradox-free does not lead to bad properties, it does not give any information on the *efficiency* of such policies. For example, an allocation α that would be defined as returning 1/1000 of the optimum of the aggregation operator sum to all users would obviously be Braess-paradox-free but is very inefficient. This calls for more quantitative characterization of efficiency.

9.2 Fairness

9.3 Price of Anarchy

9.4 Multi-Objective Related Measures

9.4.1 Considered Problem

Consider a traveler who goes from Paris to Lyon. He can either take the train (quick but expensive) or a bus (cheap but slow). What choice should he make?

This problem belongs to a family called *multi-criteria* or *multi-objective* optimizations. In these problems, we do not aim at maximizing one criteria, but simultaneously optimizing severals.

Mathematically speaking, if $(u_i)_{i \in 1..I}$ is a set of I optimization criteria, the problem consists of finding x that maximizes $(u_1(x), u_2(x), ..., u_I(x))$ subject to the system constraints.

Obviously, if the criteria are conflicting (like price and duration of a trip), no solution x can optimize simultaneously the different criteria.

Example 1: Scheduling

Consider a set of n tasks and m identical machines. The size of a task is the time it needs to be processed on a machine. The scheduling problem consists in finding an assignment (and an ordering) of tasks to machines such that it maximizes some criteria.

Typical criteria found in the literature are:

- The *makespan*, that is the time of completion of the last task in the system. This represents the point in time where all tasks have been processed.
- The *average completion time* is the average time stamp at which tasks have finished their execution in the system.
- The *average response time* or *average flow* is the average difference between the completion time of a task and its arrival time in the system.
- The *average stretch* or *slowdown* is the average response time per unit size of a task (i.e. divided by the processing time).

Among the various scheduling policies that were proposed in the literature, we can mention the list scheduling family. In a list schedule, the different tasks are ranked according to some criteria in a waiting list. Then, each time one of the machines of the system becomes idle, the task ranked first in the waiting list is executed in the machine and removed from the list. Two basic ordering policies were proposed, namely, the increasing size order and decreasing size order. The associated schedules are the SPT (Shortest Processing Time) and the LPT (Longest Processing Time).

These two policies are, in a sense opposite, and hence exhibit good performance according to different criteria: the SPT is efficient for the average completion time while the LPT exhibit good performance when considering the makespan.

Example 2: Networking

Usually, the considered quality-of-service (QoS) performance measure is the *throughput*, that is to say the average amount of information that is transmitted per second. This is particularly relevant in the case of large files, where the overall transmission time is proportional to the throughput.

Yet, in real-time environments, like audio or video communications, the system response time is crucial, and a large *delay* is not acceptable. Because of the long distance the signal has to go through, satellite communications suffer from a noticeable delay.

In streaming video, another performance measure, the *jitter* is of particular interest. The jitter is the variation of delay. When watching a video, a sudden increase of the delay might cause the buffer to be empty and the video might stop.

These different performance measures are often conflicting. For instance, jitter is usually caused by retransmission of lost packets. By adding redundancy to the transmission, a lower loss probability can be achieved, at a cost of a lower throughput.

Depending on the type of considered application, one or several of these performance measure can be appropriate to represent the quality of a connection.

9.4.2 Different Approaches

Different approaches were proposed in the literature. We review below a few popular ones.

Aggregating the Criteria

The first option is to aggregate the different criteria into a global one:

Definition 12 (Index function or Aggregation Operator)**.** *An index function f is a function of the form:* $f : \begin{cases} \mathcal{U}(\mathcal{S}) \subset \mathbb{R}_+^n & \mapsto \mathbb{R} \\ u = (u_1, u_2, ..., u_I) & \to f(u) \end{cases}$

Definition 13 (Index-optimizing)**.** *A policy function α is said to be f-*optimizing *if for all $\mathcal{U}(\mathcal{S})$, $f(\alpha(\mathcal{U}(\mathcal{S}))) = \sup_{u \in \mathcal{U}(\mathcal{S})} f(u)$ [49].*

In spite of the old saying, "you can't compare apples and oranges", this is typically done in a number of works where the utility associated to an allocation is a gain (some performance evaluation measure) minus a cost function. The "gain" can be the throughput, the probability of successful transmission, while the "cost" could be the energy consumption (mobile devices) or the price paid (to obtain a service).

Common indexes are:

- **Arithmetic mean**: $\sum_i u_i$.
- **Minimum**: $\min_i u_i$.
- **Maximum**: $\max_i u_i$.
- **Geometric Mean**: $\prod_i u_i$.
- **Harmonic Mean**: $\frac{1}{\sum_i 1/u_i}$.
- **Quasi-arithmetic Mean**: $f^{-1}(\frac{1}{n}\sum_{i=1}^n f(u_i))$ where f is a strictly monotone continuous function on $[0, +\infty]$. The particular case where f is defined by $f : x \to x^\delta$ has been widely studied [50]. The five previous index are particular case of this index for particular values of δ (respectively, $1, -\infty, +\infty, 0$ and -1).

- **Jain**: $\frac{(\sum u_i)^2}{n \sum u_i^2}$ (see [51]).

- **Ordered Weighted Averaging**: $OWA(u_1, \ldots, u_n) = \sum_i w_i.u_{\sigma(i)}$ where σ is a permutation such that $u_{\sigma(1)} \leq u_{\sigma(2)} \leq \cdots \leq u_{\sigma(n)}$.

All these indexes are continuous, however, some of them are not strictly monotone. We will see in the following that this has important repercussions (Section 9.1.2).

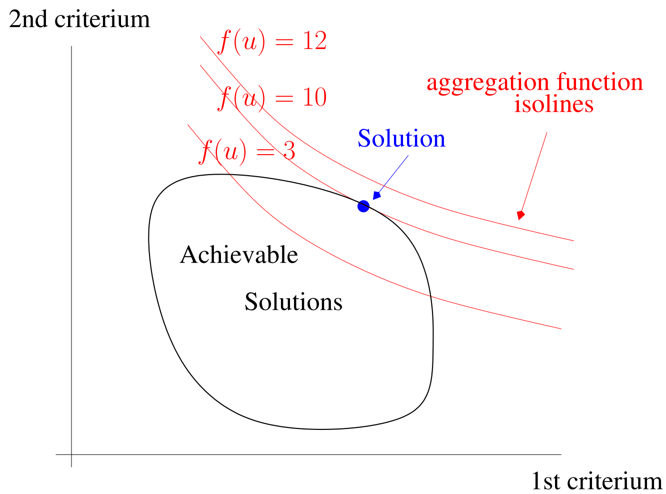


Figure 9.7: Example of aggregation optimization: A case with two criteria. The set of "achievable solutions" is the set of points $(\overline{u_1}, \overline{u_2})$ such that exists an possible x such that $u_1(x) = \overline{u_1}$ and $u_2(x) = \overline{u_2}$. The curves in red represent the isolines of the aggregation function f for different values. The optimum value is obtained at the blue point, corresponding to a value of 10 for the aggregation function.

Optimizing One, Bounding the Others

Another option is to optimize one of the criteria, while bounding the others to some acceptable levels. In this case, the optimization problem becomes:
For some $j \in \{1..I\}$, and some $(\varepsilon_i)_{i \in 1..I, i \neq j}$, Find x that maximizes $u_j(x)$ such that $\forall i \neq j, u_i(x) \geq \varepsilon_i$.

For instance, in a video communication, one might want to (lower) bound the delay by 0.2 seconds, while maximizing the throughput so as to use a compression rate with quality loss as low as possible.

Hierarchical Optimization

Let us rank the different criteria. Let σ a permutation in $1..I$, and an order $\prec$ on the criteria: $u_{\sigma(1)} \prec u_{\sigma(2)} \ldots$. Among the different solutions that maximize $u_{\sigma(1)}$, consider those that also maximize $u_{\sigma(2)}$, and among them, whose that maximize $u_{\sigma(3)}$ and so on.

This approach is suited in systems where there is a main criterion whose optimum solution is not unique. Then, the other criteria, serve as tie-breaking rule.

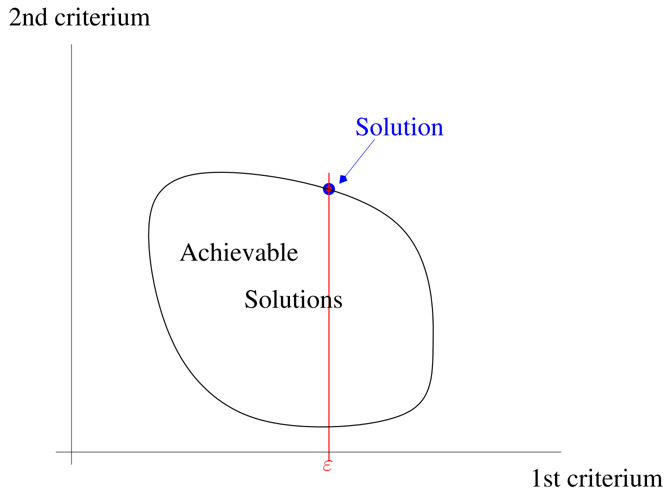


Figure 9.8: Optimizing and Bounding. The set of "achievable solutions" is the set of points $(\overline{u_1}, \overline{u_2})$ such that exists an possible x such that $u_1(x) = \overline{u_1}$ and $u_2(x) = \overline{u_2}$. The solution point x maximizes $u_2(x)$ constraint to $u_1(x) \geq \varepsilon_1$.

Zenith Optimization

Consider these I optimization problems:

$$\begin{cases} \overline{x_1} = \text{argmax } u_1(x) \\ \overline{x_2} = \text{argmax } u_2(x) \\ \dots \\ \overline{x_1} = \text{argmax } u_I(x) \end{cases}$$

The point $u_1(\overline{x_1}), u_2(\overline{x_2}), ..., u_I(\overline{x_I})$ represents the ideal point where all criteria would be maximized. Of course such a point does not exists (life is not perfect). It is called the *zenith.*

Then, an algorithm is said $(A_1, A_2, ..., A_I)$-optimal if, in the worse case (worse possible scenario, in terms of topology for instance), it computes a point y such that $\forall i, u_i(\overline{x})/u_i(y) \leq 1/A_i$.

Suppose that S_{Cmax} is a scheduling policy that is ρ_{Cmax}-optimal for the makespan and S_{MinSum} is a scheduling policy that is ρ_{MinSum}-optimal for the sum of the completion time (recall that these are objectives to be minimized). Let α be a parameter (given). Let $l = \alpha C_{max}^{S_{cmax}}$. If tasks are adequately split into two sets and scheduled according to S_{Cmax} or S_{MinSum} depending on the set they belong to, it can be shown that the obtained schedule is a $((1+\alpha)\rho_{Cmax}, (1+1/\alpha)\rho_{MinSum})$-approximation of the zenith solution of MaxAndSum (Stein and Wein, [52]).

9.4.3 Links and Use with Game Theory

As mentioned earlier, *Game Theory* deals with multiple user, each of them optimizing a single valued function. Meanwhile, in *multi-objective* approaches, a single user aim at optimizing a multiple valued function.

The different approaches used in multi-objective optimization have reciprocal applications in game theory

- Some *aggregation operators* are used to define fair sharing, mostly in resource allocation based problems (cf lecture on bargaining and fairness).

- *optimizing one, bounding the others* is useful when different types of users interact (for real time applications, which require a fixed amount of bandwidth vs elastic applications)

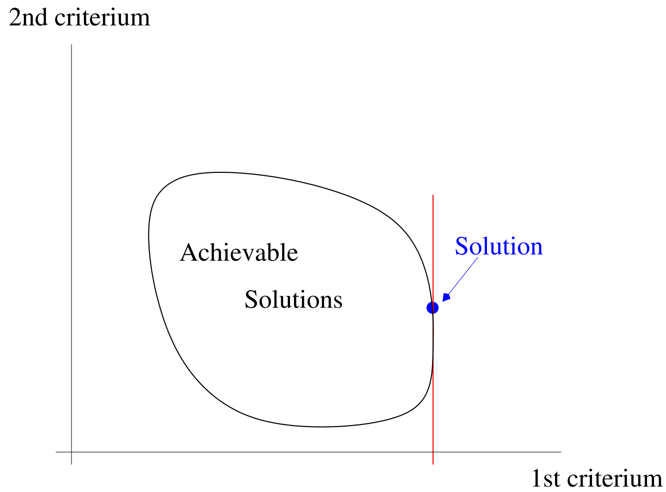


Figure 9.9: Hierarchical Optimization. The set of "achievable solutions" is the set of points $(\overline{u_1}, \overline{u_2})$ such that exists an possible x such that $u_1(x) = \overline{u_1}$ and $u_2(x) = \overline{u_2}$. The solution point x maximizes $u_2(x)$. As more than one point satisfies this property, the point maximizing $u_2(x)$ is chosen in this set.

- *hierarchical optimization*
- *Zenith optimization* is used for measuring inefficiencies of points (see Section 9.5).

9.5 Selfish Degradation Factor

How to measure the efficiency of a given policy is still an open question. We discuss in this section the popular *Price of Anarchy* [53] and the *Selfish Degradation Factor*, a more topological point of view and explain how it relates to the notion of ϵ-approximation [54].

9.5.1 Price of Anarchy and Index-Optimizing Based Metrics

Index-optimizing based metrics are easy to compute, continuous and generally conserve Pareto-superiority (under some mild conditions). It is thus natural to select an index f and to try to compare an allocation to the optimal one for f. Papadimitriou [53] introduced the now popular measure "price of anarchy" that we will study in this section.

For a given index f, let us consider $\alpha^{(f)}$ a f-optimizing policy function. We define the inefficiency $I_f(\beta, U)$ of the allocation $\beta(U)$ for f as

$$\begin{aligned} I_f(\beta, U) &= \frac{f(\alpha^{(f)}(U))}{f(\beta(U))} \geq 1 \\ &= \max_{u \in U} \frac{f(u)}{f(\beta(U))}. \end{aligned} \tag{9.1}$$

Papadimitriou focuses on the arithmetic mean Σ defined by $\Sigma(u_1, \dots, u_k) = \sum_{k=1}^{K} u_k$. The price of anarchy ϕ_Σ is thus defined as the largest inefficiency:

$$\phi_\Sigma(\beta) = \sup_{U \in \mathcal{U}} I_f(\beta, U) = \sup_{U \in \mathcal{U}} \frac{\sum_k \alpha^{(\Sigma)}(U)_k}{\sum_k \beta(U)_k}$$

In other words, $\phi_\Sigma(\beta)$ is the *approximation ratio* of β for the objective function Σ. This measure is very popular and rather easy to understand. However, we will see that it may not reflect what people have in mind when speaking about "price of anarchy".

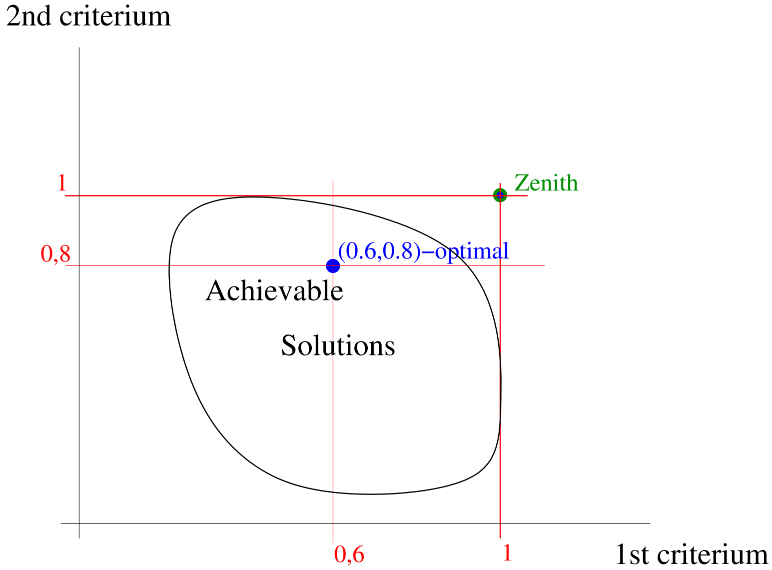


Figure 9.10: Zenith Optimization. The set of "achievable solutions" is the set of points $(\overline{u_1}, \overline{u_2})$ such that exists an possible x such that $u_1(x) = \overline{u_1}$ and $u_2(x) = \overline{u_2}$. The zenith

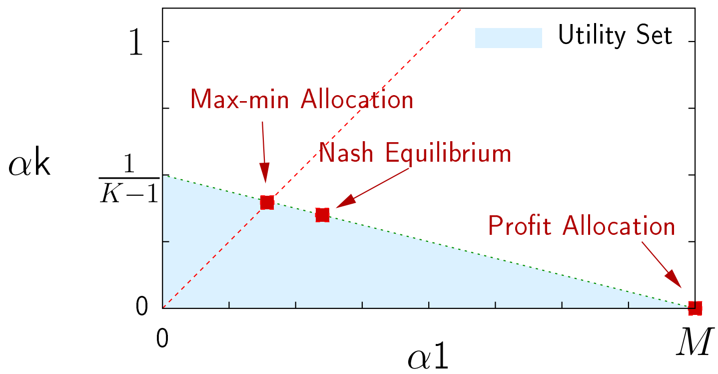


Figure 9.11: Utility set and allocations for $S_{M,N}$ ($N = 3$,$M = 2$), with $u_2 = \cdots = u_N$.

Consider the utility set $S_{M,N} = \{u \in \mathbb{R}_+^N | u_1/M + \sum_{k=1}^N u_k \leq 1\}$ depicted in Fig 9.11. As the roles of the u_k, $k \geq 2$ are symmetric, we can freely assume that $u_2 = \cdots = u_N$ for metrical index-optimizing policies.

Remark 8. *This example was taken from the master-slave scheduling problem of [47].*

It is then easy to compute the following index optimizing allocation:

- $\alpha^{(\Sigma)}(S_{M,N}) = (M, 0, \ldots, 0)$ corresponds to the allocation optimizing the average utility (sum aggregator);
- $\alpha^{(\min)}(S_{M,N}) = \left(\frac{1}{N-1+1/M}, \ldots, \frac{1}{N-1+1/M}\right)$ corresponds to the symmetric Pareto point (called max-min fair allocation [55], see lecture on bargaining and fairness);
- $\alpha^{(\Pi)}(S_{M,N}) = \left(\frac{M}{N}, \frac{1}{N}, \ldots, \frac{1}{N}\right)$ corresponds to the optimum of the product of utilities, also called proportionally fair allocation (product aggregator, see course on bargaining)

Note that, $\alpha^{(\Sigma)}$, $\alpha^{(\min)}$, and $\alpha^{(\Pi)}$ are Pareto optimal by definition. One can easily compute the price of anarchy of the Product aggregator optimizer:

$$I_\Sigma(\alpha^{(\Pi)}, S_{M,N}) = \frac{M}{\frac{M}{N} + \frac{N-1}{N}} \xrightarrow[M\to\infty]{} N.$$

The price of anarchy is therefore unbounded. However, the fact that this allocation is Pareto-optimal and has interesting properties of fairness (it corresponds to a Nash Bargaining Solution [55], see lecture on bargaining) questions the relevance of the *price of anarchy* notion as a Pareto efficiency measure.

Likewise, the inefficiency of the max-min fair allocation is equivalent to M for large values of M (as opposed to K for the non-cooperative equilibrium). It can hence be unbounded even for bounded number of applications and machines. This seems even more surprising as such points generally result from complex cooperation and are hence Pareto optimal. These remarks raise once more the question of the measure of Pareto inefficiency.

These are due to the fact that a policy optimizing an index f is always non-monotone for a distinct index g (from Theorem 14). Hence any policy (including Pareto optimal ones) optimizing a distinct index from the arithmetic mean will experience a bad price of anarchy. Note that the previous problems are not specific to the efficiency measure arithmetic mean. The same kind of behavior can be exhibited when using the min or the product of the throughputs for instance.

That is why Pareto inefficiency should be measured as the *distance* to the Pareto border and not to a specific point.

9.5.2 Selfishness Degradation Factor

Definition

The distance from $\beta(U)$ to the closure of the Pareto set $\overline{\mathcal{P}}(U)$ in the log-space is equal to:

$$d_\infty(\log(\beta(U), \log(\overline{\mathcal{P}}(U))) = \min_{u\in\overline{\mathcal{P}}(u)} \max_k |\log(\beta(U)_k) - \log(u_k)|$$

Therefore, we can define

$$\begin{aligned}\mathcal{I}(\beta, U) &= \exp(d_\infty(\log(\beta(U), \log(\overline{\mathcal{P}}(U))) \\ &= \min_{u\in\overline{\mathcal{P}}(u)} \max_k \max\left(\frac{\beta(U)_k}{u_k}, \frac{u_k}{\beta(U)_k}\right)\end{aligned} \tag{9.2}$$

Topological interpretation and connection with ε-approximation

To quantify the degradation of Braess-like Paradoxes (the degree of Paradox), Kameda [56] introduced the Pareto-comparison of α and β as $\rho(\alpha, \beta) = \min_k \frac{\alpha_k}{\beta_k}$. Therefore, α is strictly superior to β iff $\rho(\alpha, \beta) > 1$. Intuitively ρ represents the performance degradation between α and β.

Indeed, what we are interested in is in fact some kind of distance of a point to the Pareto set. As researchers are used to look at factors when evaluating the performance of an algorithm, this distance to the Pareto set should be measured in the log space. As we have seen in the previous section, the inefficiency measure for the selfishness degradation factor is closely related to the distance to the Pareto set. More precisely, we prove that being close to the Pareto set implies a small measure of inefficiency. However, the converse is true only when the utility set has some particular properties.

Let us recall the classical expansion definition:

$$X \oplus a = \{y | d(x, y) \leq a, \text{ for some } x \in X\}$$

This definition can be easily expanded as:

$$\begin{aligned}X \otimes a &= \exp(\log(X) \oplus \log(a)) \\ &= \{y | \exp(d(\log(x), \log(y)) \leq a \text{ for some } x \in X\}\end{aligned}$$

Definition 14 (ϵ-approximation)**.** *[54] defines an ϵ-approximation of $\overline{\mathcal{P}}(U)$ as a set of points S such that for all $u \in U$ there exists some $s \in S$ such that $\forall k : u_k \leq (1+\epsilon)s_k$.*

Remark 9. *Note that a $(A, A, .., A)$-(Zenith) optimization is a A-approximation of the Zenith.*

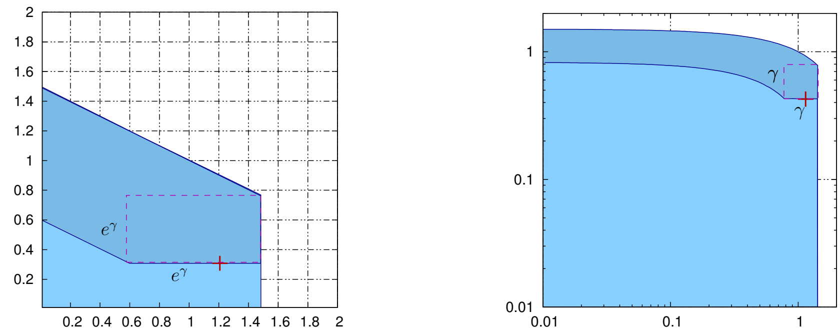
With the previous notations, it is easy to see that:

Theorem 16. *$S \subseteq U$ is an ϵ-approximation of $\overline{\mathcal{P}}(U)$ iff $\overline{\mathcal{P}}(U) \subseteq S \otimes \exp(\epsilon)$.*

Figure 9.12(b) depicts the expansion of $\log(\overline{\mathcal{P}}(U))$ by ϵ so that it contains $\log(\beta(U))$. It is easy to show that:

Lemma 4. *$\mathcal{I}(\beta, U) \leq \exp(\epsilon) \Leftrightarrow \beta(U) \in \overline{\mathcal{P}}(U) \otimes \exp(\epsilon)$.*

In other words, $\mathcal{I}(\beta, U) \leq \exp(\epsilon)$ iff $\beta(U)$ is no farther than ϵ from $\overline{\mathcal{P}}(U)$ in the log space.



(a) U and $\beta(U)$ in the original space

(b) U and $\beta(U)$ in the log-space: $\log(\beta(U)) \in \log(\overline{\mathcal{P}}(U)) \oplus \gamma$

Figure 9.12: Distance to the Pareto set

Conclusion

When comparing the definitions of I_Σ and I, the latest may seem harder to compute as it relies on $\P(U)$. However, what we are interested in is measuring the distance to the Pareto set and no index-based inefficiency measure can reflect this distance. Then can only reflect a particular property of the allocation such as fairness. Note that in mono-criteria situations, it is natural to compare a solution to an intractable optimal solution, generally using approximations or lower bounds. Therefore, similar approaches should be used in multi-criteria settings to compute I. This inefficiency measure is thus a natural extension of the classical mono-criteria performance ratio.

A system (e.g., queuing network, transportation network, load-balancing, ...) that would be such that the Nash equilibria are always Pareto optimal would have a selfishness degradation factor equal to one. The selfishness degradation factor may however be unbounded on systems where non-cooperative equilibria are particularly inefficient. The relevance of this definition is corroborated by the fact that ϵ-approximations of Pareto-sets defined by Yannakakis and Papadimitriou [54] have a degradation factor of $\exp(\epsilon) \simeq 1 + \epsilon$.

9.6 Examples: Schedulling

We present two games that fall into this categories. The first is the problem of when to arrive at a bank. The second is of when to retry to make a phone call. We then briefly mention a third related game that also involves retransmission decisions and that is related to access to a radio channel.

9.6.1 When to arrive at the bank?

The aim of this example is to illustrate that a very common situation that we frequently face is in fact a queueing game, where the decision is of when to queue.

We consider the following scenario. A bank opens between 9h00 to 12h00. All customers that arrive before 12h00 are served that day. A random number X of customers wish to get a service at a given day. Service time are i.i.d. with exponential distribution. The order of service is FCFS (First Come First Served). Each customer wishes to minimize her own expected waiting time, i.e. the expected time elapsed from her arrival at the bank till she gets served.

This problem has been studied in [57]. The authors show that there exists a symmetric equilibrium distribution F of the arrival time with support $[T_0, 12h00]$ for all players, with $T_0 < 9h00$, such that if all customers follow F, then no individual has an incentive to deviate from using the distribution F.

We finally note that if we eliminate the FCFS regime among those who arrive before the bank opens, then $T_0 = 9h00$. This could reduce average waiting time!

9.6.2 When to retry to make a phone call?

We are all used to obtaining occasionally a busy signal when attempting to make phone calls. This signal is either due to the fact that the destination is currently busy with another phone call or to congestion that may cause calls to be blocked. A person whose call is blocked may typically retry calling. The goal of the example, taken from [58, 59], is to analyze the individual's choice of time between retrials.

The model: Calls arrive according to a Poisson process with average rate λ. Service rates are i.i.d. with mean τ and finite variance σ^2. Let $S^2 := \tau^2 + \sigma^2$ and $\rho := \lambda\tau$. Between retrials calls are said to be "in orbit". Times between retrials of the ith call in orbit are exponentially distributed with expected value of $1/\theta_i$.

We note that exponentially distributed times are frequently used in models in telecommunications, and in particular in telephony models due to much accumulated experimental results. In particular, it is known that the duration of phone calls is well modeled by the exponential distribution. The exponential distribution is also attractive mathematically since it leads to simple Markovian models due to its memoryless property (i.e. the property $P(X > t + s | X \geq t) = P(X > s)$ for any $t, s > 0$).

We assume that each retrial costs c, and the waiting time costs w per time unit.

We begin by presenting the performance of the socially optimal policy [60].

The socially optimal policy. The expected time in orbit when all retrials use the same parameter θ is

$$W = \frac{\rho}{1-\rho}\left(\frac{1}{\theta} + \frac{S^2}{2\tau}\right),$$

so the average cost per call is

$$(w + c\theta)W =$$

$$\frac{\rho}{1-\rho}\left(\frac{cS^2}{2\tau}\theta + \frac{w}{\theta}\right) + \frac{\rho}{1-\rho}\left(\frac{wS^2}{2\tau} + c\right).$$

This is minimized at

$$\theta^* = \frac{\sqrt{2w\tau/c}}{S}.$$

A remarkable observation is that this rate is independent of the arrival rate!

If θ^* is used, the two terms that depend on θ turn out to be equal: the <u>waiting</u> cost and the <u>retrial</u> costs coincide.

The game In the game case, the authors of [60] compute $g(\theta, \gamma)$, which is the expected waiting time of an individual who retries at rate γ while all the others use retrial rate θ. This allows them to obtain the equilibrium rate:

$$\theta_e = \frac{w\rho + \sqrt{w^2\rho^2 + 16w\tau c(1-\rho)(2-\rho)/S^2}}{4c(1-\rho)}.$$

We observe that θ_e monotone increases to infinity as λ increases (to $1/\sigma$).

Conclusions. Here are some conclusions from this example.

1. The ratio between the cost at equilibrium and the globally optimal cost tends to infinity. Thus, the noncooperative nature of the game leads to high inefficiency.
2. The equilibrium retrial rate is larger than the globally optimal retrial rate. They tend to coincide as $\rho \to 0$.
3. Both equilibrium and optimal retrial rates are monotone decreasing in the variance of the service times.
4. There is a unique equilibrium.

9.6.3 Access to a radio channel: when to retransmit

Phone calls are not the only example where we find retransmissions. Another example, typical to cellular communications as well as satellite communications, is in the access of packets to a common radio channel. If more than one terminal attempt to transmit a packet at the same time then a collision occurs and all terminals involved in the collision have to retransmit their packet later. A terminal wishes typically to maximize the throughput (i.e. the expected average number of successful transmissions). A further cost may be incurred per each transmission attempt (it could represent the energy consumption). The problem has various formulations as stochastic games, in which each terminal controls the time till it attempts a retransmission of a packet that has collided, or the probability of its retransmission at a given time slot [**?**, 61, 62]. In [**?**], a comparison between a cooperative solution and a noncooperative solution is presented. It is shown that as the arrival rate of packets increases, the cooperative solution tends to reduce the retransmission probabilities; retransmissions are thus delayed so as to avoid more collisions. Due to this behavior, the throughput was shown to increase with the arrival rate of packets. In the noncooperative case, a single symmetric equilibrium is obtained in which, in contrast, the retransmission probabilities tend to increase with the arrival rate of packets. Thus, as the arrival rate increases, terminals become more aggressive, and the system throughput decreases. For large arrival rates it decreases to zero.

Chapter 10

Computing Equilibria

10.1 Zero-Sum matrix games

We introduced games and equilibria first through some properties related to a symmetric 2×2 matrix game with 2 players and then through their standarnd definitions. We show now that the definition implies this structure in a more general setting.

Theorem 17. *Consider a 2×2 matrix game with 2 players. The game need not be symmetric. If the equilibrium strategy of a player is not pure then that player is indifferent between his actions.*

Proof. Let (p, r) be an equilibrium. Assume that p is not pure. p maximizes $U^1(q, r)$ over q where

$$U^1(q, r) = qU^1(T, r) + (1 - q)U^1(B, r).$$

Since p is an interior point, it is obtained at the q for which the derivative of $U^1(q, r)$ equals zero. The latter gives

$$U^1(T, r) = U^1(B, r).$$

■

Theorem 18. *Consider a 2×2 matrix game with 2 players where player 1 has the utility matrix A (as in Figure 10.1). Assume that no player has a dominating strategy. Then there is a unique mixed equilibrium (p, r) where*

$$r = \frac{A_{22} - A_{12}}{A_{11} - A_{12} - A_{21} + A_{22}} \tag{10.1}$$

If the game is symmetric then $p = r$ and the utility of each player at equilibrium is

$$U^1(r, r) = \frac{A_{11}A_{22} - A_{12}A_{21}}{A_{11} - A_{12} - A_{21} + A_{22}} \tag{10.2}$$

Proof. We have

$$U^1(T, r) = rA_{11} + (1 - r)A_{12}, \quad U^1(B, r) = rA_{21} + (1 - r)A_{22}.$$

Equating them, we get (10.1).

	Player 2	H	D
Player 1	H	A_{11}	A_{12}
	D	A_{21}	A_{22}

Figure 10.1: Utility for player 1

Next we show that r is indeed a probability measure. We note that the enumerator of (10.1) is non-zero. Indeed, if $A_{22} = A_{12}$ then player 1 has a dominating strategy: it is 1 if $A_{11} \geq A_{21}$ and is 2 if $A_{11} \leq A_{21}$.

We show that the denominator is non-zero. Assume $A_{11} \geq A_{21}$. Then $A_{12} < A_{22}$ otherwize 1 would be a dominant strategy for player 1. Hence $A_{11} - A_{21} - A_{12} + A_{22} < 0$ Assume $A_{11} \leq A_{21}$. Then $A_{12} > A_{22}$. Hence $A_{11} - A_{21} - A_{12} + A_{22} > 0$.

We conclude that the denominator in the expression for (10.1) for r is non-negative and that indeed $0 < r < 0$.

Since $U^1(B,r) = U^1(T,r)$, p is indeed optimal when the other player uses r. We can similarly derive the optimal r when the other player uses p. Thus (p,r) where r is given in (10.1) are in equilibrium.

Substituting r in the expression $rA_{11} + (1-r)A_{12}$ for $U^1(1,r)$ gives (10.2). ∎

10.2 Computing Wardrop Equilibria

Basic Wardrop model

Recalll that at Wardrop equilibrium the flows satisfy the following (Section 6.1).

$$h_{wr}(c_{wr} - \pi_w) = 0, r \in R_w, w \in W, \tag{10.3}$$

$$c_{wr} - \pi_w \geq 0, r \in R_w, w \in W, \tag{10.4}$$

$$\sum_{r \in R_w} h_{wr} = d_w, w \in W \tag{10.5}$$

where $h_{wr}$is the flow on route $r \in R_w$, the set of routes joining node pair $w \in W$,the set of origin-destination node-node pairs. The cost or delay on that route is c_{wr}, and π_w is the minimum cost on any route joining node pair w. The demand for service between the node pair w is denoted d_w. Define the graph as $G = (\mathcal{N}, \mathcal{A})$. Then, adding non-negativity restrictions $h_{wr} \geq 0$ and $\pi_w \geq 0$, the resulting system of equalities and inequalities can be seen as the Karush-Kuhn-Tucker (KKT) optimality conditions of the following optimization problem, known as the Beckmann transformation (Beckmann, 1956).

$$\min f(x) = \sum_{l \in A} \int_0^{x_l} t_l(x_l) dx = \sum_{l \in A} \int_0^{\sum_{i \in N} x_{il}} t_l(x) dx$$

subject to

$$\sum_{r \in R_w} h_{wr} = d_w, \quad w \in W, \tag{10.6}$$

$$\sum_{w \in W} \sum_{r \in R_w} h_{wr} \delta^l_{wr} = x_l, l \in A, \tag{10.7}$$

$$x_l \geq 0, l \in A, \tag{10.8}$$

where x_{il} is the flow of users from OD pair i on link l, x_l is $\sum_{i \in N} x_{il}$ and δ^l_{wr} is a $0-1$ indicator function that takes the value 1 when link l is present on route $r \in R_w$.

10.2.1 Wardrop equilibrium and Potential games

The fact that Wardrop equilibrium can be obtained using an equivalent optimization problem with a single player having some cost $f(x)$ is a feature common to a whole class of games known as potential games. This class of games was formally introduced by Monderer and Shapley (1996) for the case of finitely many players. It was extended in Sandholm (2000) to the case of population games, which includes the setting of Wardrop equilibrium.

In developing the concept of potential games, game theorists seem not to have been aware of the huge literature on road traffic equilibria starting from Wardrop (1952) and Beckmann (1956). Monderer and Shapley write in Monderer and Shapley (1996): "To our knowledge, the first to use potential functions for games in strategic form was Rosenthal (1973)". Interestingly enough, this reference (see also Rosental, 1973b) includes a discrete version of Wardrop equilibrium with finitely many players, called "congestion games".

The original definition of a potential for a game is as follows. Introduce the following N-player game $G = (N; (S^i)_{i\in N}; (\eta^i)_{i\in N})$ where S^i is the action set of player i, $S = \times_{i\in N} S^i$ and $\eta^i(s)$ is the payoff for player i when the multistrategy $s \in S$ is used. For $s \in S$, let $(s|t^i)$ denote the multistrategy in which player i uses t^i instead of s^i and other players $j \neq i$ use s^j. A potential for the game is defined in Monderer and Shapley (1996) as a real valued function P on S s.t. for each i, every $s \in S$ and every $t^i \in S^i$, $P(s|t^i) - P(s) = \eta^i(s|t^i) - \eta^i(s)$. Existence and uniqueness of equilibria of potential games in that setting has been established in Monderer and Shapley (1996) and Neyman (1997)[1].

An adaptation of this definition is needed for population games, see Chap. 3 of Patriksson (1994) and Sandholm (2000), in which there are N classes of populations of "infinitesimal" players, where the "mass" of players of type i is given by some constants d_i. Let $\alpha(j,t)$ be the fraction of members of population type j that use action $t \in S^i$. A multistrategy is the collection $\alpha = (\alpha(j,t))$. We assume that the payoff η^i for a player of class i is a function of his own action as well as of the multistrategy α. Let S^i be the set of actions available to a player of population i, $i = 1, ..., N$. We say that α^* is an equilibrium if for any i, any $s \in S^i$ and any $t \in S^i$ such that $\alpha^*(i,t) > 0$, $\eta^i(t;\alpha^*) \geq \eta^i(s;\alpha^*)$. Equivalently, letting $f = -\eta$, we say that flow α^* is in equilibrium if the following variational inequality problem (VIP) holds for all $\alpha \in S$: $f(\alpha^*)(\alpha - \alpha^*) \geq 0$.

We then define P to be a potential for the population game if for each s, the vector of payoffs $\eta(s)$ is the gradient of $P(s)$. Under mild conditions on the payoff functions and strategy sets, one can thus establish the existence and uniqueness of equilibria in potential population games. (See Chap. 3 of Patriksson, 1994.)

10.2.2 Commodity–link-variable models

In telecommunication network models, the variable of interest is often the commodity-link flow, x_{il}, rather than the total link flow, x_l. In this section, we consider a typical model based on commodity-link flows as the decision variable, and through it, derive some relations between Nash and Wardrop equilibria.

We have seen that the Wardrop equilibrium can be computed using an equivalent convex optimization problem (related to the potential). We show below that reminiscence of this potential appears also in problems of Nash equilibria with finitely many users: all equilibria satisfying some conditions on the commodity-link flow (condition (10.13) below) have the same total link flow that can be obtained from some convex optimization problem.

In the network routing problem, the decision variables x_{il} are restricted by the non-negativity constraints for each link l and player i: $x_{il} \geq 0$ and by the conservation constraints for each player i and each node v:

$$r_v^i + \sum_{j\in In(v)} x_{ij} = \sum_{j\in Out(v)} x_{ij} \tag{10.9}$$

where $r_v^i = d_i$ if v is the source node for player i, $r_v^i = -d_i$ if v is its destination node, and $r_v^i = 0$ otherwise; $In(v)$ and $Out(v)$ are respectively all ingoing and outgoing links of node v. (d_i is the total demand of player i).

A typical commodity-link cost function, for the ith commodity on link l, is given by

$$f_{il}(x_{il}) = x_{il} t_l \left(\sum_{j\in\mathcal{I}_l} x_{jl} \right), \tag{10.10}$$

[1] Uniqueness is in fact established among the class of correlated equilibria, of which Nash equilibria is a subset.

where $\mathcal{I}_l$ is the set of commodities (players) using link l. Note that this model resembles the system optimal model defined in Wardrop's second principle. The Lagrangian with respect to the conservation of flow constraints is

$$L_i = \sum_{l \in \mathcal{A}} x_{il} t_l \left(\sum_{j \in \mathcal{I}_l} x_{il} \right) + \sum_{v \in \mathcal{N}} \pi_{i,v} \left(r_v^i + \sum_{j \in In(v)} x_{ij} - \sum_{j \in Out(v)} x_{ij} \right),$$

for each player i. Thus a vector x with nonnegative components satisfying (10.9) for all i and v is an equilibrium if and only if the following Karush-Kuhn-Tucker (KKT) condition holds:

Let $x_{uv} = \sum_{i \in \mathcal{A}} x_{i,uv}$, where $x_{i,uv}$ is the flow of users from OD pair i on the link defined by node pair u, v. There exist Lagrange multipliers $\pi_{i,u}$ for all nodes u and all players, i, such that for each pair of nodes u, v connected by a directed link (u, v),

$$t_{uv}(x_{uv}) + x_{i,uv} \frac{\partial t_{uv}(x_{uv})}{\partial x_{uv}} \geq \pi_{i,u} - \pi_{i,v},$$

with equality if $x_{i,uv} > 0$.

Define $\pi_u = \sum_i \pi_{i,u}$. Taking the sum over all players we get the following necessary conditions for x to be an equilibrium for each link (u, v):

$$It_{uv}(x_{uv}) + x_{uv} \frac{\partial t_{uv}(x_{uv})}{\partial x_{uv}} \geq \pi_u - \pi_v, \tag{10.11}$$

with equality if $x_{i,uv} > 0$ for all i, where I is the total number of players.

Assume that all players have the same source and destination, and let d be the sum of commodity demands. Then (10.11) are the KKT conditions for optimality of the vector $\{x_l\}$ with nonnegative components satisfying the conservation of flow constraints in the routing problem (single commodity) where the cost to be minimized is given by

$$\sum_{l \in A} x_l t_l(x_l) + (I - 1) \int_0^{x_l} t_l(y) dy, \tag{10.12}$$

and where the total demand to be shipped from the common source to the common destination is d; in particular, (10.11) holds with equality if $x_{uv} > 0$. Assume further that t_l are strictly convex, or more generally that expression (10.12) is strictly convex. Then this problem has a unique solution in total link flows, which we denote (x_l^*).

Now, let $\{x_{il}\}$ be a Nash equilibrium for the original problem having costs (10.10) with the property:

A1. Whenever $x_{il} > 0$ for some i and l then $x_{jl} > 0$ for all players j. (10.13)

A1 describes a property of the equilibrium: if (at equilibrium) one player sends positive flow through a link, then so do all other players. Under assumption A1, it follows that for all l, $\sum_i x_{il} = x_l^*$. Note however, that this is not true in general if A1 does not hold, since x_l^* need not be expressible as the sum over I of some nonnegative x_{il} that satisfy (10.9).

Remark The above is an alternative proof to the one in Orda, Rom and Shimkin (1993) of the uniqueness of the total link flows at all Nash equilibria satisfying A1.

Taking the limit in (10.12) as the number of players $I \to \infty$, the second term in (10.12) dominates, and by continuity of the functions and compactness of the feasible set, we observe that both the objective function and the solution approach that of the Wardrop equilibrium.

10.2.3 General models and variational inequalities

The basic equilibrium model imposes a number of simplifications on the model of the traffic flow phenomenon, and in particular, on the travel time, or impedance, functions.

Most notably, for the potential function to exist, the travel time function, t_l, defined for each link of the network, $l \in A$, must be integrable. The most common way for this to occur is that the travel time on a link l depends only upon the flow present on the link l, that is, $t_l(x) = t_l(x_l)$. This simplification, in the traffic context, means that interactions between different traffic streams at junctions cannot be modeled within this paradigm (even if we use a virtual link to model the node), since then, the travel time on a link l that reaches the junction is a function of flows on some or all links meeting link l at the junction, that is, $t_l(x) = t_l(x_1, \ldots, x_l, \ldots, x_m)$ (and is not just a function of the sum of flows). The simplification also imposes that only a single class of users is modeled, since multiple classes of users would interact on each link, resulting once again in multivariate link travel time functions. In short, when the identities (in terms of multiple user classes, or the multiple links they use) is needed in the cost function of a single link, then the single-class model is no longer applicable.

When the link travel time functions are multivariate, it is usually the case that no potential that can be obtained by integrating the travel time functions as in the basic model. Examples of multivariate link cost functions can be found in the literature on modeling signalizes junctions on a road network Heydecker (1983). Other examples can be found in the modeling of multimodal networks, such as networks on which buses and cars share the road space or trucks and light vehicles, since each traffic class effects the traffic differently, and each class has its own travel time function, depending on all classes present on the link. Some characteristics of this type of multivariate cost functions can be found in Toint and Wynter (1995).

Although the potential function approach cannot be used to describe the multivariate equilibrium, the Wardrop equilibrium conditions are valid regardless of whether the cost functions are univariate or multivariate, and they can be expressed for both types of cost functions in a compact variational framework.

In telecommunication network planning, link impedance functions can often be quite complex, due to the underlying probabilistic phenomena as well as the interacting cost components of delay, packet loss, jitter, etc... A typical form of the commodity-link cost functions (see, for example, Orda Rom and Shimkin, 1993) is

$$t_{il}(x_{il}, x_l) = x_{il} t_l(x_l) = \frac{x_{il}}{C_l - x_l}, \tag{10.14}$$

where, as before, $x_l = \sum_{i=1..I} x_{il}$ is the flow of all classes i on link l. The constant C_l is the capacity of link l. The user classes in this case correspond to commodities, or origin-destination demands. For more justification on this type of delay models, see Baskett et al. (1975), Kameda and Zhang (1995). In some simple settings, such as a network of parallel arcs, when the sum of the demands of all classes is less than link capacity, or on some one-commodity networks, Orda, Rom and Shimkin (1993) show uniqueness of the Nash equilibrium. Also under "diagonal strict convexity", the authors show uniqueness of the Nash equilibrium, yet this condition only holds under quite restrictive conditions on the cost functions or on the topology (Altman et al., 2002). In Orda, Rom and Shimkin (1993) and Altman and Kameda (2001), uniqueness of the link class flows is shown for costs of the form (10.14) for general networks under assumption A1 (see eq. (10.13).

We can analyze the *Wardrop* equilibrium for this system easily by expressing it as the solution $\{x_i^*\}$ of the following variational inequality:

$$t_i(x^*)^T (x_i^* - x_i) \leq 0, \tag{10.15}$$

for all feasible class-flow vectors, x_i, where, as above, the vector $x = \sum_{i=1..I} x_i$ and $t_i = t$ for all classes $i = 1..I$. Since the Jacobian of the mapping t is clearly singular ($\partial t_i(x)/\partial x_j = \partial t_j(x)/\partial x_i$ for all i, j), we recover the nonuniqueness of the equilibrium (in the variable x_i) that was observed by Orda, Rom and Shimkin (1993) to occur on general networks.

We can also contrast the Wardrop and Nash equilibria through this example. The Nash equilibrium x_i^* satisfies

$$\sum_{l \in R_i} x_{il}^* t(x_{il}^*, x_{\neq il}^*) \leq \sum_{l \in R_i} x_{il} t(x_{il}, x_{\neq il}^*)$$

for each user class $i = 1..I$, where the index $\neq i$ includes all classes not equal to i. Rewriting, we obtain that $\sum_{l\in R_i} x^*_{il} t(x^*_{il}, x^*_{\neq il}) - \sum_{l\in R_i} x_{il} t(x_{il}, x^*_{\neq il}) \leq 0$ for each $i \in I$ and therefore when

$$t(x^*_{il}, x^*_{\neq il}) - t(x_{il}, x^*_{\neq il}) = 0, \tag{10.16}$$

the above reduces to $\sum_{l\in R_i}(x^*_{il} - x_{il}) t(x^*_{il}, x^*_{\neq il}) \leq 0$, for each $i \in I$, which is equivalent to the (Wardrop) VIP with cost operator t and classes given by users $i \in I$. Indeed, (10.16) occurs precisely when the influence of an additional user on the cost, t, is 0.

Another example of general costs in telecommunications can be found in Altman, El-Azouzi, and Abramov (2002), where the network model includes dropping of calls if capacity is exceeded. The cost criterion for each user class i is the probability that his message is rejected along its path, given by

$$B_i(x) = 1 - \frac{\sum_{\mathbf{m}\in S_i} \prod_{j=1}^{I} (x_j^{m_j}/m_j!)}{\sum_{\mathbf{m}\in S} \prod_{j=1}^{I} (x_j^{m_j}/m_j!)}.$$

S is the set of feasible "states"; S_i is the set of states for which another call of user i can still be accepted (without violating capacity constraints), $\mathbf{m}$ is the system state whose jth component m_j is the number of class j calls in the system. Hence the number of terms in the cost function depends upon the configuration and capacity constraints of the network. (The larger the number of feasible paths for a user class, the more terms present.) When each class, which can be represented by an origin-destination pair, seeks a Wardrop equilibrium, the corresponding variational inequality is to find x^*_i such that $B_i(x^*)^T(x^*_i - x_i) \leq 0$, for all x_i, for all classes i. Even in the simplest topology of parallel links, it has been shown in Altman, El-Azouzi, and Abramov (2002), that there may exist several equilibria with different total link flows.

Depending on the form of monotonicity satisfied by the cost operator in the variational inequality, one can choose convergent algorithms for its solution, as well as determine whether or not the solution will be unique. Weaker forms of monotonicity for which the mathematical properties and a number of convergent algorithms are known include pseudo-monotonicity and strong nested monotonicity. (See Marcotte and Wynter, 2001, Cohen and Chaplais, 1988 ,for the latter.)

10.2.4 Additive versus non-additive models

The most widely studied performance measure investigated to date in transportation, computer and telecommunication networks has been the expected delay. In transportation networks, this cost metric leads naturally to models in which the route costs are additive functions of link costs along the route, that is $c_r(x) = \sum_{l\in R} t_l(x_l)$.

In telecommunication networks, exogenous arrivals of jobs or of packets are modeled as Poisson processes. Delays at links are modeled by infinite buffer queues with i.i.d. service times, independent of the interarrival times. In the particular case in which the service time at a queue is exponentially distributed, then whenever the input process has a Poisson distribution so will the output stream. This makes the modeling of service times through exponential distributions quite appealing, and makes the cost along a path of tandem queues additive. However, the expected delay turns out to be additive over constituent links in a general topology under a much more general setting, known as BCMP networks (named after its authors Baskett et al., 1975).

Indeed, as long as the exogenous arrivals have Poisson distributions and under fairly general assumptions on the service order and service distribution, the expected delay over each link is given by the expected service time divided by (1-ρ), where ρ is the product of total average traffic flow at the queue and the expected service time. This general framework also allows for the modeling of multiclass systems, i.e. where different traffic classes require different expected service times at a queue, see Kameda and Zhang (1995).

Other more general types of separable additive cost functions have been used in telecommunication networks which can represent physical link costs due to congestion pricing, see e.g. Orda, Rom and Shimkin (1993).

However, in both transportation and telecommunication networks, equilibrium models in which path costs are not the sum of link costs do arise.

In the transport sector, when environmental concerns are taken into account, such as the pollution associated with trips, non-additive terms arise. Path costs due to tolls or public transport costs are generally non-additive as well, since they are calculated over entire paths, and cannot be decomposed into a sum of the costs on component links.

In Gabriel and Bernstein (1997) and Bernstein and L. Wynter (2000), the authors discussed properties of a bicriteria equilibrium problem in which both time delay and prices are modeled on the links. Non-additivity arises from the nonlinear valuation of the tradeoff between time and money, known as a nonlinear *value of time*. For example, users are willing to pay more (or less) per minute for longer trips than for shorter trips. That is, a route cost function may be expressed as $c_r(x,p) = V\left(\sum_{l\in R_i} t_l(x_l)\right) + \sum_{l\in R_i} p_l$, where $V : \Re_+^m \mapsto \Re_+$ is the nonlinear value of time function.

In telecommunication networks, many important performance measures are neither additive nor separable. The first example is that of loss probabilities when the network contains finite buffer queues. We note that in this case there is no flow conservation at the nodes[2] This model has been dealt with by light traffic approximations which are additive and separable, see Dinan, Awduche and Jabbarie (2000), Jiménez (2001).

Another performance measure (already mentioned in the previous subsection) that has been studied in the context of network equilibrium is that of rejection probabilities. The network consists of resources at each link, and requests for connections between a source and a destination. The resources are limited, and a connection can only be established if there are sufficiently many resources along each link of a route between the source and destination. For the case where connections arrive according to a Poisson process and where calls last for an exponentially distributed duration (these assumptions model telephone networks well), simple expressions for rejection probabilities (i.e. the probability that an arriving call will find the line busy) are available. These expressions are neither separable nor additive. Such networks have been studied in Altman, El-Azouzi, and Abramov (2002), Bean, Kelly, and Taylor, (1997).

When route costs are no longer the sum of constituent link costs, algorithms for solving the network equilibrium problems must be modified; indeed, underlying most algorithms for the network equilibrium problem is a shortest path search, and standard searches all suppose additive path costs. One algorithm for the non-additive route cost model can be found in Gabriel and Bernstein (2000) while another is provided in the article by M. Patriksson [XXX].

[2]Note that flow conservation fails even in the case of infinite queues when one considers multicast applications in which packets are duplicated at some nodes, see Boulogne and Altman (2002).

Chapter 11

Routing Games: Applications

11.1 The Braess paradox

The service providers or the network administrator may often be faced with decisions related to upgrading of the network. For example, where should one add capacity? Where should one add new links?

A frequently-used heuristic approach for upgrading a network is through *bottleneck analysis*, where a system bottleneck is defined as "a resource or service facility whose capacity seriously limits the performance of the entire system" (see p. 13 of [?]). Bottleneck analysis consists of adding capacity to identified bottlenecks until they cease to be bottlenecks. In a non-cooperative framework, however, this heuristic approach may have devastating effects; adding capacity to a link (and in particular, to a bottleneck link) may cause delays of all users to increase; in an economic context in which users pay the service provider, this may further cause a decrease in the revenues of the provider. This problem was identified by Braess [63] in the transportation context, and has become known as the *Braess paradox*. See also [?], [?]. The Braess paradox has been studied as well in the context of queuing networks [?], [64], [65], [66], [?].

In the latter references both queuing delay as well as rejection probabilities were considered as performance measures. The impact of the Braess paradox on the bottleneck link in a queuing context as well as the paradoxical impact on the service provider have been studied in [?]. In all the above references, the paradoxical behavior occurs in models in which the number of users is infinitely large and the equilibrium concept is that of Wardrop equilibrium, see [?].

It has been shown, however, in [?], [67], that the problem may occur also in models involving a finite number of players (e.g. service providers) for which the Nash framework is used. The Braess paradox has further been identified and studied in the context of distributed computing [?], [?], [?] where arrivals of jobs may be routed and performed on different processors. Interestingly, in those applications, the paradox often does not occur in the context of Wardrop equilibria; see [?].

In [?] (see also [?]), it was shown that the decrease in performance due to the Braess paradox can be arbitrarily larger than the best possible network performance, but the authors showed also that the performance decrease is no more than that which occurs if twice as much traffic is routed. The result was extended and elaborated upon in more recent papers by the same authors. In [?], a comment on the results of [?] was made in which it is shown that if TCP or other congestion control is used, rather than agents choosing their own transmission rates, then the Braess phenomenon is reduced considerably. Indeed, this conclusion can be reached intuitively by considering (as is well known in the study of transportation equilibria) that the system optimal equilibrium model (in which the sum of all delays are minimized) does not exhibit the Braess paradox; congestion control serves to force transmission rates to such a system optimal operating point.

An updated list of references on the Braess paradox is kept in Braess' home page at http://homepage.ruhr-uni-bochum.de/Dietrich.Braess/#paradox

11.2 Architecting equilibria and network upgrade

The Braess paradox illustrates that the network designer, the service provider, or, more generally, whoever is responsible for setting the network topology and link capacities, should take into consideration the reaction of (non-cooperative) users to her or his decisions. Some guidelines for upgrading networks in light of this have been proposed in [?], [?], [?], [?], [67], so as to avoid the Braess paradox, or so as to obtain a better performance. Another approach to dealing with the Braess paradox is to answer the question of which link in a network should be upgraded; see, for example, [?] who computes the gradient of the performance with respect to link capacities.

A more ambitious aim is to drive the equilibrium to a socially optimal solution. In [?] this is carried out under the assumption that a central manager of the network has some small amount of his or her own flow to be shipped in the network. It is then shown that the manager's routing decision concerning his own flow can be taken in a way so that the equilibrium corresponding to the remaining flows attain a socially optimal solution.

11.3 Analyzing the original Braess paradox

We introduce in Figure 11.1 the topology of the original Braess paradox. The total demand is L.

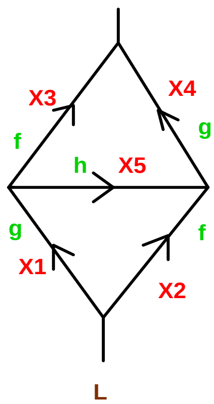


Figure 11.1: The topology of the original Braess paradox

The Figure provides for every link, its corresponding cost density as well as the amount of flow in that link. We shall show how the structure of the cost functions chosen simplifies the computations of the equilibrium.

We notice that the costs have been chosen such that the following properties hold:

- P1. link 3 and link 2 have the same cost function associated to them.
- P2. link 1 and 4 have the same cost.
- P3. the cost function f (of links 2 and 3) is linear
- P4. $f + g$ is strictly increasing.

These properties simplify a lot the computations of the equilibrium, as we shall see next.

Case 1: x_5 is not used at equilibrium. Then flow conservation implies that $x_1 = x_3$ and $x_2 = x_4$. Hence

$$g(x_1) + f(x_1) = g(x_2) + f(x_2)$$

Property P4 then implies that $x_1 = x_2$. The conservation of flow at the source then imply that

$$x_1 = x_2 = x_3 = x_4 = \frac{L}{2}.$$

Case 2: x_5 is also used at equilibrium. Assume that the cost is such that both route 1-3, as well as 2-4 are used at equilibrium. Then according to Wardrop's principle

$$g(x_1) + f(x_3) = f(x_2) + g(x_4). \tag{11.1}$$

We now express x_5:

$$x_5 = x_1 - x_3 = -x_2 + x_4.$$

The two equalities follow from the conservation of flows at the node to the left and to the right of link 5.

As f is linear then this equation implies

$$f(x_1) - f(x_3) = -f(x_2) + f(x_4).$$

Summing with (11.1), we get

$$f(x_1) + g(x_1) = f(x_4) + g(x_4).$$

As $f + g$ is strictly increasing then this equation implis $x_1 = x_4$. Hence also $x_2 = x_3$ (by the previous equation).

Now, the conservation of flows at the source node implies that $x_2 = L - x_1$. Hence

$$x_5 = x_1 - x_3 = x_1 - x_2 = 2x_1 - L.$$

As route 1-5-4 is also used then

$$g(x_1) + h(x_5) = f(x_2).$$

We conclude that

$$g(x_1) + h(2x_1 - L) = f(L - x_1).$$

This gives x_1.

In particular, if g, h, f are all linear with the form

$$f(x) = f_a x + f_b, \quad g(x) = g_a x + g_b, \quad h(x) = a_1 x + h_b$$

Then

$$x_1 = \frac{-g_b + h_a L - h_b + f_a L + f_b}{g_a + 2h_a + f_a}$$

which as we saw, equals also x_4, and

$$x_3 = x_2 = L - x_1 = \frac{g_b + h_a L + h_b - f_b + g_a L}{g_a + 2h_a + f_a}$$

In the original reference of Braess we have $L = 6$, $f(x) = 50 + x$, $g(x) = 10x$. Thus $f_a = 1, f_b = 50, g_a = 10, g_b = 0$. If link 5 is absent (or if h_2 is very large) then we get Then $x_1 = x_2 = 3$,

$$D_{13} = D_{24} = 83.$$

Take $h = 10 + x$ so that $h_a = 1$ and $h_b = 10$. We observe that $x_1 = x_2 = x_3 = x_4 = 3$ is no more an equilibrium. Indeed, With $x_1 = x_3 = 3$, the cost of the unused path is $D_{154} = 70$ which is strictly smaller than that of the path that is used: $D_{13} = 83$.

The Wardrop equilibrium turns out to be $x_1 = 4 = x_4$, $x_2 = L - x_1 = 2 = x_3$, $x_5 = x_1 - x_2 = 2$, At equilibrium, the delays over all paths equal 92.

Chapter 12

Improving the Quality of Equilibria

12.1 Setting up Appropriate Prices

12.1.1 Shadow Prices

12.1.2 Pigou Prices

12.1.3 Repercussion Prices

12.2 Coordinating Players: the Role of the Correlator

The concept of Correlated equilibrium is due to R. J. Aumann [68]. The following game is taken from that reference:

	action 1.a	action 1.b
action 2.i	$2, 1$	$0, 0$
action 2.ii	$0, 0$	$1, 2$

Table 12.1: A matrix game [Aumann]

The game has two pure Nash equilibria:
(1.a , 2.i) with the values (2,1), and (1.b , 2.ii) with values (1,2).

The randomized strategy (1/3,2/3) for player 1 and (2/3,1/3) for player 2 is (the unique) symmetric equilibrium. The corresponding values are

$$(2/9) \times 1 + (2/9) \times 2 = 2/3 \quad \text{for each player}$$

Suppose that an arbitrator **suggests** to both either $(1.a,\ 2.i)$ (w.p. $1/2$) or $(1.b,\ 2.ii)$ (w.p. $1/2$). If the players follow the advice they obtain in expectation $3/2$ each. Cannot be obtained without correlation.

This is an example of a *correlatead game*. The game is still non-cooperative: no "binding contract". Here, the correlated equilibrium does not dominate the pure Nash equilibria. There are cases where it dominates all other Nash equilibria. The coordinator does not have to know anything about the game: all it has to do is flip a coin and send signals. Correlation is needed for **coordination**.

Correlation is useful not just in games but also in team problems (a common objective).

Chapter 13

General equilibrium: Dealing with Constraints

13.1 Routing games and constraints

We represent a network as a graph with a set of nodes and directional links. A link has a **cost** (delay) per flow unit, that depends on the total flow through it.

Consider K classes of users, each class i has a fixed demand ϕ_i to ship from a source s^i to a destination d^i node. The strategy of a user is given by a vector describing the amount of flow to be shipped over each link of the network.

The cost of a path is the sum of the link costs.

We consider the following constraints:
(i) non-negative flows and
(ii) conservation constraints: for each node and class, the sum of entring class flow equals the sum of its leaving flows.

These are **Orthogonal constraints:** The constraints of class i do not depend on the strategies of other classes.

13.1.1 Routing: Common capacity constraints

We add to the model of the previous subsectionn side constraints. Each link ℓ has a **capacity constraint**: the total flow over it cannot exceed C_ℓ.

We say that link ℓ has a **Common Capacity Constraint** (CCC) for users in a group $M(\ell)$ if violation of the capacity of link ℓ for some $i \in M(\ell)$ implies its violation $\forall j \in M(\ell)$.

ℓ has a CCC for some $M(\ell)$ iff for any multistrategy satisfying the link constraints, $\forall i \in M(\ell)$ send positive traffic on ℓ.

Interpretation: If the capacity of link ℓ is exceeded then all those using it suffer (delay or losses). Users not using the link are not affected.

CCC are non-orthogonal. The constraints of class i depend on the strategies of other classes.

Next consider the case of N **classes**, as described in Fig. 13.1.1.

User i has a **dedicated direct path** to the destination (link Si–Di) and an **alternative path** Si–Bi–B2–Di that shares a common bus B1–B2 with the others.

Example: assume the demand ϕ^i of player i is of i units and that the capacity of each direct links Si–Di is k. As k decreases, the number of users that have to ship flow over the common bus increases.

Hence **the bus is a CCC** for all users in a group $M(S_k)$ where S_k **decreases in** k. It is given by $S_k = \{N, N-1, ..., N-k\}$.

2 Users Example (fig 13.1):
Each user has a direct dedicated path and a second alternative path that traverses a common bus.

Let the demands of class 1 and 2 be given by $\phi^1 = \phi^2 = 4$, $C_l = 3, \forall l$.
Then $\ell = (S1 - S2)$ has a CCC for $M(\ell) = \{1, 2\}$.

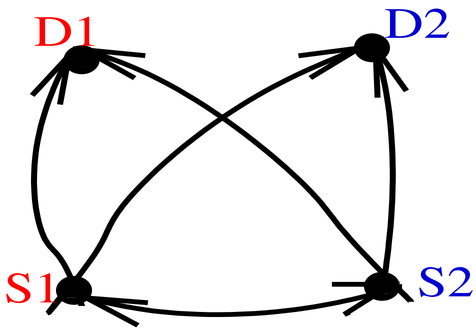
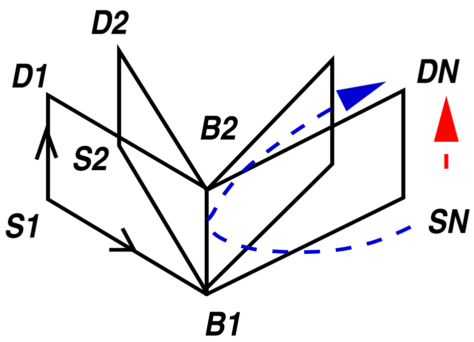


Figure 13.1: 2 users Example



13.2 Rosen coupled constraints: non-uniqueness fo value

13.2.1 Example: matrix game

		Player 2: L	Player 2: R
Player 1	T	0 (2)	1 (0)
	B	-1 (0)	0 (0)

- Player 1 (2) chooses T or B. (L or R, respectively).
- The strategies of player 1 and 2 are the probabilities written as row vectors: $x = (x(T), x(B))$ and $y = (y(T), y(B))$, respectively.
- There are two matrices: U - utilities and D - constraints. The entries of U (of D) are in black (in red, resp.).
- Player I maximizes the expected outcome xUy' and player II minimzes it. (y' is the transposed of y).
- As in Rosen, *the constraint is common to both players*:

$$xDy' \leq \rho. \tag{13.1}$$

where ρ is some constant taken to be 0 in this Example.

		Player 2	
		L	R
Player 1	T	0 (2)	1 (0)
	B	-1 (0)	0 (0)

		Player 2	
		L	R
Player 1	T	0 (2)	1 (0)
	B	-1 (0)	0 (0)

Consider $x = (1, 0)$ (choose T with probability 1). In order for the constraint to hold, player II has to play $y = (0, 1)$ (choose R with probability 1). Hence

$$\max_x \min_y xUy' = U(T, R) = 1.$$

Next assume Player II chooses $y = 1$. To meat the constraint, Player I has to play B with probability 1. Hence

$$\min_y \max_x xUy' = U(B, L) = -1.$$

We conclude that a value does not exist. Moreover, we obtain the surprising unusual inequality

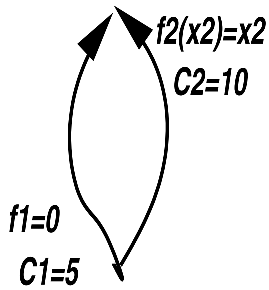
$$\max_x \min_y xUy' > \min_y \max_x xUy'.$$

13.2.2 Networking game Example: parallel links

x_k^i is the amount of flow that player i sends over link k.

$f_k(x_k)$:= cost per unit flow of link k, where $x_k = x_k^1 + x_k^2$ is the total flow on link k.

Let $f_1 = 0$, $f_2(x_l) = x_l$, $C_1 = 5, C_2 = 10$, $\phi^1 = \phi^2 = 3$.
Player 1 maximizes $J(\mathbf{x}) = \mathbf{x_2^1}(\mathbf{x_2^1} + \mathbf{x_2^2})$ and player 2 minimizes it.



If player 1 plays first then its dominating strategy - ship all his demand to link 1: $x_1^1 = 3, x_2^1 = 0$ and player, $J(\mathbf{x}) = 0$, so that $\min_{x^1} \sup_{x^2} J(\mathbf{x}) = 0$.

The same holds for player 2, who has a dominating strategy of sending all her flow to link 1. Hence

$$\sup_{x^2} \min_{x^1} J(\mathbf{x}) = 1.$$

We thus get again the surprising inequality: $\min_{x^1} \sup_{x^2} J(\mathbf{x}) < \sup_{x^2} \min_{x^1} J(\mathbf{x})$..

13.3 Properties of zero-sum games with Rosen's constraints

Consider a zero-sum game with a set $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2$ of multi-strategies.

$\mathcal{S}_i(z)$:= non-empty set of strategies available to user i when the other player $j \neq i$ plays z.

A multi-strategy (x, y) is feasible if $x \in \mathcal{S}(y)$ and $y \in \mathcal{S}(x)$.

Let the **unrestricted game** be the game with no constraints

Theorem 19. *Assume that in the unrestricted game a value exists, i.e.*

$$\sup_{x \in \mathcal{S}_1} \inf_{y \in \mathcal{S}_2} U(x,y) = \inf_{y \in \mathcal{S}_2} \sup_{x \in \mathcal{S}_1} U(x,y) \tag{13.2}$$

Then the original constrained game satisfies

$$\sup_{x \in \mathcal{S}_1} \inf_{y \in \mathcal{S}_2(x)} U(x,y) \geq \inf_{y \in \mathcal{S}_2} \sup_{x \in \mathcal{S}_1(y)} U(x,y) \tag{13.3}$$

13.3.1 Remarks

In constrained zero-sum games, the maximization and minimization by players 1 and 2 are restricted to **feasible multi-strategies**.
We can relax this: in the left hand side or (13.3), the maximization of player 1 is over all S_1; player 2 takes care that the constraints of player 1 are saatisfied.
A symmetric argument holds for the right hand side.

(13.3) shows that the behavior observed for Rosen's constraints, where the "upper-value" is smaller than the "lower-value", is typical for constrained games of Rosen's type.

(13.3) holds also in constrained games without Rosen's structure of common constraints. But it does not have any more a useful interpretation since

1. the "lower-value" (in the left hand side) is not restricted any more to multi-strategies that are feasible for player 1, and

2. the "upper-value" (in the right hand side) is not restricted any more to multi-strategies that are feasible for player 2.

13.4 Zero-sum games with general constraints

We now raise the question of whether a value exists when the constraints do not satisfy the framework of Rosen

We begin considerinig the case where only player 1 has constraints. These depend on the stgrategies of both players.

13.4.1 Formulation of the problem

Let S_1 be the set of strategies available to player 1. Player 2's strategies when player 1 uses strategy x are given by the set $S_2(x) = \{y : D(x,y) \leq \rho\}$.

If player 2 knows x then it can play any strategy in $S_2(x)$. If it does not know x it is obliged to play a strategy y that will guarantee that $\sup_{x \in \mathcal{S}_1} D(x,y) \leq \rho$.

Let $G_2 = \{y : \forall x,\ D(x,y) \leq \rho\}$. The general problem we consider is of determining the relation between the left and the right hand side of the following:

$$\sup_{x \in S_1} \inf_{y \in S_2(x)} U(x,y) \begin{matrix} ? \\ > \\ = \\ < \end{matrix} \inf_{y \in G_2} \sup_{x \in S_1} U(x,y)$$

The following holds for any $x^* \in \mathcal{S}_1$, and $y^* \in G_2$:

$$\inf_{y \in S_2(x^*)} U(x^*,y) \leq \sup_{x \in S_1} \inf_{y \in S_2(x)} U(x,y) \leq \inf_{y \in G_2} \sup_{x \in S_1} U(x,y) \leq \sup_{x \in S_1} U(x,y^*)$$

(x^*, y^*) is said to be a value of the game where $x^* \in \mathcal{S}_1$, $y^* \in G_2$, if

$$\inf_{y \in S_2(x^*)} U(x^*,y) = \sup_{x \in S_1} U(x,y^*)$$

If it holds, then we conclude that

$$\inf_{y\in S_2(x^*)} U(x^*,y) = \sup_{x\in S_1}\inf_{y\in S_2(x)} U(x,y) = \inf_{y\in G_2}\sup_{x\in S_1} U(x,y) = \sup_{x\in S_1} U(x,y^*)$$

Note that the Rosen's joint coupled constraints are not special cases of this framework, so our previous Counterexamples do not allow us to conclude that a value does not exist here.

Symetric definitions hold when player 1 has constraints. Then if a value exists, i.e.

$$\inf_{y\in S_2} U(x^*,y) = \sup_{x\in S_1(y^*)} U(x,y^*)$$

for some $x^* \in G_1$ and $y^* \in S_2$, then

$$\inf_{y\in S_2} U(x^*,y) = \sup_{x\in G_1}\inf_{y\in S_2} U(x,y) = \inf_{y\in S_2}\sup_{x\in S_1(y)} U(x,y) = \inf_{y\in S_1(x^*)} U(x,y^*).$$

13.4.2 Constrained Matrix games

Example 4. Zero-sum matrix game with constraints on player 1's strategies

the constraint restricts only player 1 and not player 2. Set $\rho = 1/2$.

		Player 2	
		L	R
Player 1	T	1 (1)	0 (0)
	B	0 (0)	2 (1)

G_1 is the singleton $(1/2, 1/2)$. Hence $\sup_{x\in G_1}\inf_{y\in S_2} U(x,y) = 1/2$, obtained when player 2 uses action R w.p.1.

Thus Player 1 can guarantee to receive a payoff of at least 1/2 if he plays first.

If Player 2 plays first, he can guarantee that the payoff of Player 1 would not exceeed 2/3.

We conclude that the following inequality holds:

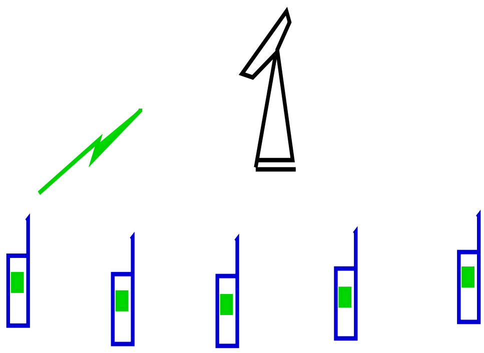
$$\sup_{x\in G_1}\inf_{y\in S_2} U(x,y) < \inf_{y\in S_2}\sup_{x\in S_1(y)} U(x,y).$$

∎

The inequality is the opposit than the one we had in Rosen's setting

13.4.3 Power control: general constraints

N mobile terminals. Each minimizes its transmission power P_i.



The transmissions of all mobiles are received at a base station (BS).

The received power of mobile i is $h_i P_i$. ξ is the power of the thermal noise at BS.

The quality of the call of mobile i is determined by the SINR (Signal to Initerference and Noise Ratio):

$$SINR_i(\mathbf{P}) = \frac{h_i P_i}{\sum_{j \neq i} h_j P_j + \xi}, \qquad \mathbf{P} = (P_1, ..., P_N).$$

Each mobile i seeks to **Minimize** P_i subject to $SINR_i(\mathbf{P}) \geq \rho_i$.
The consstraint of a user depends on strategies of other users.
The contstrainsts are non-orthogonal and non common.

13.5 Lagrangian Multipliers and Duality Theory

Chapter 14

Stochastic games: applications - Where to queue?

14.1 The gas station game

The example we present here illustrates the dynamic routing choices between two paths. When a routing decision is made, the decision maker knows the congestion state of only one of the routes; the congestion state in the second route is unknown to the decision maker. The problem originates from the context of two gas stations on a highway [69]. A driver arriving at the first station sees the amount of other cars already queued there and has to decide whether to join that queue, or to proceed to the next gas station. The state of the next gas station (i.e. the number of cars there) is not available when making the decision. The situation is illustrated in Fig. 14.1. The exact mathematical solution of the model was obtained in [70] and we describe it below.

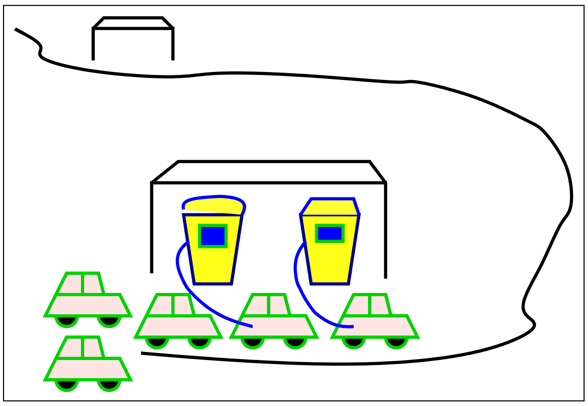


Figure 14.1: The gas station problem

This problem has natural applications in telecommunication networks: when making routing decisions for packets in a network, the state in a down stream node may become available after a considerable delay, which makes that information irrelevant when taking the routing decisions.

Although the precise congestion state of the second route is unknown, its probability distribution, which depends on the routing policy, can be computed by the router.

We assume that the times that corresponds to the arrivals instants of individuals is a Poisson process with rate λ. Each arrival is a player, so there is a countable number of players, each of whom takes one routing decision, of whether to join the first or the second service station. A player eventually leaves the system and does not affect it anymore, once it receives service.

To obtain an equilibrium, we need to compute the joint distribution of the congestion state in both routes as a function of the routing policy.

We restrict to random threshold policies (n, r):

- if the the number of packets in the first path is less than or equal to $n-1$ at the instance of an arrival, the arriving packet is sent to path 1.
- If the number is n then it is routed to path 1 with probability r.
- If the number of packets is greater than n then it is routed to path 2.

The delay in each path is modeled by a state dependent queue:

- Service time at queue i is exponentially distributed with parameter μ_i
- Global inter-arrival times are exponential i.i.d. with parameter λ.

When all arrivals use policy (n, r), the steady state distribution is obtained by solving the steady state probabilities of the continuous time Markov chain [70].

If an arrival finds i customers at queue 1, it computes

$$E_i[X_2] = E[X_2 | X_1 = i]$$

and takes a routing decision according to whether

$$T^{n,r}(i,1) := \frac{i+1}{\mu_1} \leq_? \frac{E_i[X_2]+1}{\mu_2} =: T^{n,r}(i,2).$$

To compute it, the arrival should know the policy (n, r) used by all previous arrivals.

If the decisions of the arrival as a function of i coincide with (n, r) then (n, r) is a Nash equilibrium.

The optimal response against $[g] = (n, r)$ is monotone decreasing in g. This is the **Avoid The Crowd** behavior.

Computing the conditional distributions, one can show [70] that there are parameters $(\mu_1, \mu_2, \lambda, n, r)$ for which the optimal response to (n, r) is indeed a threshold policy.

Denote

$$\rho := \frac{\lambda}{\mu_1}, \qquad s := \frac{\mu_2}{\mu_1}$$

There are other parameters for which the optimal response to (n, r) is a two-threshold policy characterized by $t^-(n, \rho, s)$ and $t^+(n, \rho, s)$ as follows.

It is optimal to route a packet to queue 2 **if** $t^-(n, \rho, s) \leq X_1 \leq t^+(n, \rho, s)$ **and to queue** 1 **otherwise.**

At the boundaries t^- and t^+ routing to queue 1 or randomizing is also optimal if $T^{n,r}(i,1) = T^{n,r}(i,2)$. For parameters in which the best response does not have a single threshold, we cannot conclude anymore what is the structure of a Nash equilibrium.

Example [70]. Consider $n = 3$, $r = 1$, $\rho = \lambda/\mu_1 = 1$ and $s = \mu_2/\mu_1 = 0.56$. We plot in Fig. 14.2 $T^{n,r}(i,1)$ and $T^{n,r}(i,2)$ for $i = 0, 1, \ldots, 4$.

Conclusions: As opposed to the example of the choice of players between a PC and a MF in Sec. **??**, we saw in this section an example where for some parameters there may be no threshold type (n, r) equilibria. The form of the Nash equilibrium in these cases remains an open problem. Moreover, even the question of existence of a Nash equilibrium is then an open question.

Yet, in the case of equal service rates in both stations, a threshold equilibrium does exist, and it turns out to have the same type of behavior as in the PC-MF game, i.e. it is unique, and the best response has a tendency of "Avoiding the Crowd" [69].

By actually analyzing (numerically) the equilibrium for equal service rates, it was noted in [69] that when players use the equilibrium strategy, then the revenue of the first station is higher

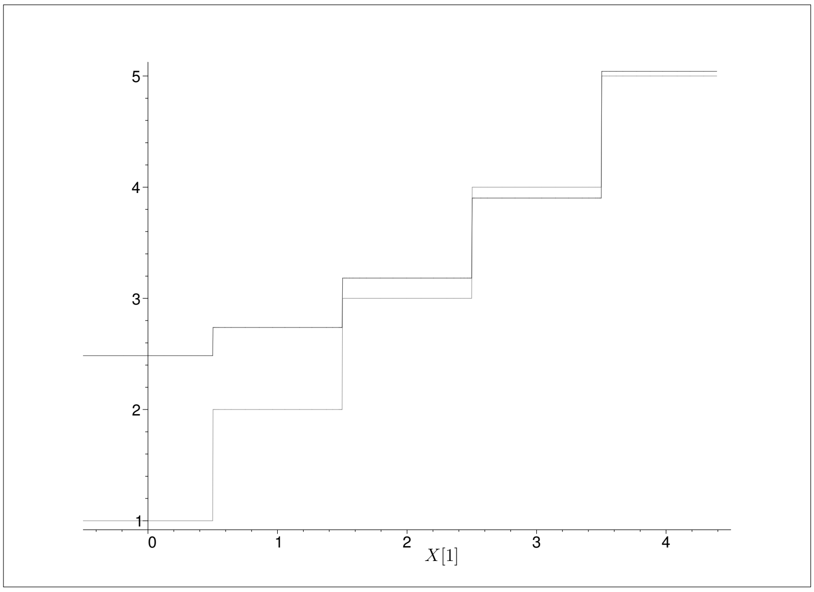


Figure 14.2: $T^{n,r}(i,1)$ and $T^{n,r}(i,2)$

than the second one. Thus the additional information that the users have on the state of the first station produces an extra profit to that station. An interesting open problem is whether the second station can increase its profits by using a different pricing than the first station, so that users will have an extra incentive to go to that station. Determining an optimal pricing is also an interesting problem.

14.2 Queues with priority

We present a second queueing problem modeled as a stochastic game with infinitely many individual players due to [71, 72]. We assume again that players arrive at the system according to a Poisson process with intensity λ, and have to take a decision of whether to join a low priority (second class) or a high priority (first class) queue, as illustrated in Fig. 14.2. There is a single server that serves both queues but gives strict priority to first class customers. Thus a customer in the second class queue gets served only when the first class queue is empty. We assume exponentially distributed service time with parameter μ and define $\rho := \lambda/\mu$.

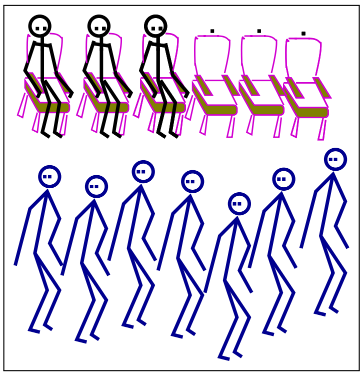


Figure 14.3: Choice between first and second class priorities

The game model is then as follows:

The actions Upon arrival, a customer (player) observes the two queues and may purchase the high priority for a payment of an amount θ, or join the low priority queue.

The state: The state is a pair of integers (i, j) corresponding to the number of customers in each queue; i is the number of high priority customers and j, the number of low priority ones.

The analysis of this problem can be considerably simplified by using the following **monotonicity property**, identified in [71]: If for some strategy adopted by everybody, it is optimal for an individual to purchase priority at (i, j), then he must purchase priority at (r, j) for $r > i$.

This implies that the problem has an effective **lower dimensional state space:** It follows that starting at $(0, 0)$ and playing optimally, there is some n such that the only reachable states are

$$(0, j),\ j \leq n, \quad \text{and } (i, n),\ i \geq 1.$$

Indeed, due to monotonicity, if at some state $(0, m)$ it is optimal not to purchase priority, it is also optimal at states $(0, i)$, for $i \leq m$. Let $n - 1$ be the largest such state. Then starting from $(0, 0)$ we go through states $(0, i)$, $i < n$, until $(0, n - 1)$ is reached. At $(0, n)$ it is optimal to purchase priority. We then move to state $(1, n)$.

The low priority queue does not decrease as long as there are high-priority customers. Due to monotonicity, it also does not increase as long as there are high-priority customers since at (i, n), $i \geq 1$ arrivals purchase priority! Therefore we remain at (i, n), as long as $i \geq 1$.

The Equilibrium. Suppose that the customers in the population, except for a given individual, adopt a common threshold policy $[g]$. Then the optimal threshold for the individual is non-decreasing in g.

This property is called **"Follow The Crowd" Behavior**

This property clearly implies *Existence* of an equilibrium, that can be obtained by a monotone best response argument.

However, it turns out that there is no uniqueness of the equilibrium! Indeed, Hassin and Haviv have shown in [72] that there may be up to

$$\left\lfloor \frac{1}{1-\rho} \right\rfloor$$

pure threshold Nash equilibria, as well as other mixed equilibria! They further present numerical examples of multiple equilibria.

We conclude that in this problem we have definitely a different behavior of the equilibria than in the previous stochastic games in which we had threshold equilibria (the PC-MF game and the gas station game).

Chapter 15

Fairness concepts

This Chaptper is based on [?]

For over two decades, the Nash bargaining solution (NBS) concept from cooperative game theory has been used in networks to share resources fairly. Due to its many appealing properties, it has recently been used for assigning bandwidth in a general topology network between applications that have linear utility functions. In this paper, we use this concept for allocating the bandwidth between applications with general concave utilities. Our framework includes in fact several other fairness criteria, such as the max-min criteria. We study the impact of concavity on the allocation and present computational methods for obtaining fair allocations in a general topology, based on a dual Lagrangian approach and on Semi-Definite Programming.

15.1 Introduction

Fair bandwidth assignment has been one of the important challenging areas of research and development in networks supporting elastic traffic. Indeed, *Max-min* fairness has been adopted by the ATM forum for the Available Bit Rate (ABR) service of ATM [?]. Although the max-min fairness has some optimality properties (Pareto optimality), it has been argued that it favors too much long[1] connections and does not make efficient use of available bandwidth. In contrast, the concept of *proportional fairness* (of the throughput assignment) has been proposed by Kelly [42, ?]. It gives rise to a more efficient solution in terms of network resources by providing more resources to shorter[1] connections.

Although the object that is shared fairly seems to be a very specific one (the throughput), it is shown in [42, ?] that in fact, the starting point for obtaining (weighted) proportional fairness of the throughput can be any general (concave) utility function per connection.

As opposed to this approach, we wish to use a fairness concept that is directly defined in terms of the users' utilities rather than of the throughputs they are assigned to. Yet, as in weighted proportional fairness, it would be desirable to express this concept as the solution of a utility maximization problem, since it makes it possible to use recent algorithms for utility maximization in networks, along with decentralized implementations [?, ?, ?].

The Nash Bargaining Solution (NBS) is a natural framework that allows us to define and design fair assignment of bandwidth between applications with different concave utilities and has already been used in networking problems [?, ?]. It is characterized by a set of axioms that are appealing in defining fairness. As already recognized in [?] and later in [?], proportional fairness agrees with the NBS if the object that is shared fairly is the throughput and if the user's minimum required rate is zero. We are interested in the NBS since it can been seen as a natural extension of the proportional fairness criterion which is probably the most popular fairness notion in network design today: it appears widely in the Internet world (indeed, it is shown in [?] that some versions

[1]The terms *long* and *short* refer here to the distance, counted in the number of hops or of saturated links for instance, as we will explain later.

the TCP congestion control protocol achieves proportional fair bandwidth sharing) as well as in wireless communications ([?]).

We use the NBS to study the fairness of an assignment where connection n has a concave utility over an interval $[MR_n, PR_n]$. It thus has a minimum rate requirement MR_n and does not need more than PR_n. Utility functions with such features have been identified in [?] for representing some real time applications such as voice and video, and in the case that $MR_n = 0$, for elastic traffic.

We briefly mention some other useful concepts for fairness or for allocating resources. One of the properties defining the Nash bargaining solution has had some criticism (an axiom stating that the solution is not affected by reducing the domain) since it implies that a player does not care how much other players have given up. Two alternative notions of fair sharing have thus been introduced with the same other properties of the Nash bargaining solution, but with a variation of the above property: the modified Thomson solution and the Raiffa-Kalai-Smorodinsky solution. A unified treatment of the Nash solution as well as of these two has been introduced in [?] for two players and extended in [?] for the multiperson case. These concepts have been applied to Internet pricing in [?]. Yet another "fair" concept for sharing resources is the Aumann-Shapley solution for cooperative game, which has desirable properties such as Pareto optimality. Haviv [?] proposes this approach to allocating congestion costs in a single node under various queueing disciplines.

We study in this paper the way the concavity of the utilities affect the bandwidth assignment according to the NBS, as well as according to a generalized version of the proportional fairness (in which the utilities that correspond to the different connections are fairly allocated).

Contribution of the paper : In previous work on Nash bargaining in networks, only linear utilities have been studied. We introduce in this paper quadratic utility functions. It is a sufficiently large class in order to represent utilities with various degrees of concavity. At the same time they are sufficiently simple for computation purposes. We indeed propose in the paper a Lagrangian (see [?]) approach as well as a reduction to Semi Definite Programming (SDP) which allows to use a large variety of open source (and other) libraries to compute efficiently the fair solution in realistic networks. We indicate the complexity of the solution by analyzing matrices that are used in the SDP formulation. We demonstrate how various fairness concepts can be handled by the proposed SDP method. We study the way the concavity of the utilities affect the bandwidth assignment according to the NBS, as well as according to a generalized version of the proportional fairness (in which the utilities that correspond to the different connections are fairly allocated). We illustrate numerically the bandwidth allocation for the Cost network [?].

Structure of the paper : After briefly introducing the notations used in the article (Sec. 15.2), we present the fair allocations in Sec.15.3 and focus on the NBS in Sec. 15.4 and its properties. We then propose a new fair family that covers for special cases the previously defined fairness criteria, including in particular the Nash, the Thomson and the Kalai-Smorondinsky solutions in Sec. 15.5 and propose a quadratic approximation for the utility of each connection, which allows us to parameterize the degree of concavity of the utility function using a single parameter. We use this approximation to further analyze the impact of concavity of utilities on the resulting assignment. Quadratic utilities can be viewed as a second order approximation (in terms of a Taylor series) of any sufficiently smooth concave utility function. We then present in Sec. 15.7 a Lagrangian approach which allows us to implement a decentralized protocol for the bandwidth allocation. We finally present in Sec. 15.8 a novel alternative approach using Semi Definite Programming (SDP) and some numerical results in Sec. 15.9

15.2 Definitions and notations

In the following, we denote by *connection* a "source-destination" pair. For a given network and set of connections, there exists a infinite set of feasible allocations. We are interested in those that are optimal (as we will define in Subsection 15.2.2) and satisfying the system's constraints (Subsection 15.2.3). The fair allocation we will present all belong to this set. Let us first summarize in Subsection 15.2.1 the notations used in this article.

15.2.1 Notations

We summarize in Table 15.1 the different notations used in the article. We suppose that to each connection n is associated a utility function f (on which we will give more insight in Section 15.4.1) and has a minimum and a maximum requirement in terms of bandwidth, that we denote MR_n and PR_n respectively.

$\|ens\|$	number of elements of set ens,	X	set of feasible allocations,
n	a connection,	$\mathcal{N}$	set of connections,
x_n	bandwidth allocated to connection n,	$\vec{x}$	vector of allocations (of size $\|\mathcal{N}\|$),
MR_n	minimal bandwidth requirements of connection n,	PR_n	maximal bandwidth required for connection n,
l	a directed link,	$\mathcal{L}$	set of links in the network,
C_l	capacity of link l,	$\vec{C}$	vector of links capacities.

Table 15.1: Notations

We also define some orders among feasible allocations (that is to say satisfying the system's constraints, as defined in Section 15.2.3). We write:

- $\vec{x} \leq \vec{y}$ if $\forall i \in 1, \dots, N, x_i \leq y_i$
- $\vec{x} \leq_{lg} \vec{y}$ if either $\vec{x} = \vec{y}$ or $\exists i, i > 0, \forall j \in 1, \dots, i-1, x_j = y_j$ and $x_i < y_i$ (lexicographic order)
- $\vec{x} \preccurlyeq \vec{y}$ if $\sigma(\vec{x}) \leq_{lg} \sigma(\vec{y})$ with $\sigma(\vec{x})$ and $\sigma(\vec{y})$ the ordered versions of $\vec{x}$ and $\vec{y}$[2].

15.2.2 Efficiency

An allocation is said to be Pareto optimal or Pareto efficient if it is impossible to increase the allocation of a connection without strictly decreasing another one. In other words, an allocation Pareto efficient is maximal in the sense of $\leq$. In the networking context, this amounts in saying that each connection goes through at least one saturated link. In general, if $|\mathcal{N}|$ is the number of connections, the set of Pareto optimal points is a subset of size $|\mathcal{N}| - 1$. The fair allocations we will present are all Pareto optimal.

15.2.3 Allocations constraints.

We briefly introduce here the constraints in the allocations vectors. They are of two types: the ones associated to the users and those due to the routing policy.

User constraints As previously mentioned, we suppose that each connection is associated to a minimum and a maximum requirement, in term of bandwidth. We therefore have a system of $|\mathcal{N}|$ independent linear inequalities $\forall n \in \mathcal{N}, MR_n \leq x_n \leq PR_n$.

Routing policy We consider the case of both fixed-routing and fractional-routing. In fixed routing, we define a matrix A of size $|\mathcal{L}| \times |\mathcal{N}|$ that specifies the links that the packets of each connection will go through:

$$A_{n,l} = \begin{cases} 1 & \text{if connection } n \text{ goes through link } l, \\ 0 & \text{otherwise.} \end{cases}$$

In some cases, we do not want to fill the links but keep some empty space, call *left-over capacity.* For instance, an operator may wish to use only up to 90% of the link in order to let some other

[2] that is to say that $\forall i$, $\sigma(\vec{x})_i$ is the i^{th} smaller element of x.

traffic join the system. Therefore we suppose that we can only use the quantity $(1-\kappa_l)C_l$ on each link l. Let κ be the vector of size $|\mathcal{L}|$ of leftover capacities. The capacity constraints can then be written:

$$A\vec{x} \leq (1-\vec{\kappa})\vec{C}. \tag{15.1}$$

In the case of fractional routing, each packet transmitted by a source can follow its own path. We denote by $\phi_n^{(u,v)}$ the flow of connection n in the link from u to v. Note that $\phi_n^{(v,u)}$ exists and is a distinct variable. V is the set of nodes and for $u \in V$, $N(u)$ is the set of nodes connected to u. Finally, $s(n)$ (respectively $d(n)$) is the source (destination) of connection n. We have 4 types of constraints for each connection n:

$$\begin{array}{ll}
x_n = \sum_{v \in N(s(n))} \phi_n^{(s(n),v)} - \sum_{u \in N(d(n))} \phi_n^{(u,s(n))} & \text{balance of flows from source } s(n), \\
x_n = \sum_{u \in N(d(n))} \phi_n^{(u,d(n))} - \sum_{v \in N(s(n))} \phi_n^{(d(n),v)} & \text{balance of flows to destination } d(n), \\
\forall u \neq \{s(n), d(n)\}, \sum_{v \in N(u)} \phi_n^{(v,u)} = \sum_{v \in N(u)} \phi_n^{(u,v)} & \text{Kirchhoff or conservation law}, \\
\forall u, v, \phi_n^{(u,v)} \geq 0. & \text{positive or null flows.}
\end{array}$$

Let us introduce $r_n^{s(n)} = -r_n^{d(n)} = 1$, and $r_n^u = 0$ for u different from $s(n)$ and $d(n)$. If we can use the fraction $(1-\kappa_l)C_l$ on each link $l = \{u, v\}$ then the capacity constraints are:

$$\forall l = \{u, v\} \in \mathcal{L}, \qquad \sum_{n \in \mathcal{N}} \phi_n^{(u,v)} \leq (1-\kappa_l)C_l$$

Note that in fact this equation applies both to (u, v) and (v, u), which means that the capacity constraints will be verified on each orientation of $\{u, v\}$.

$$\text{Finally, } \forall n \in \mathcal{N}, \begin{cases} \forall u \in V, & x_n r_n^u = \sum_{w \in N(u)} \phi_n^{(u,w)} - \sum_{w \in N(u)} \phi_n^{(w,u)}, \\ \forall \{u, v\} \in \mathcal{L}, & \phi_n^{(u,v)} \geq 0, \\ \forall l = \{u, v\} \in \mathcal{L}, & \sum_{n \in \mathcal{N}} \phi_n^{(u,v)} \leq (1-\kappa_l)C_l. \end{cases}$$

We can notice that in both cases, the constraints are linear.

15.3 Fair allocations

In this section, we review the criteria commonly used for fairly sharing resources in networks. Each of them can be found in both a simple and a weighted version. The latter was introduced in order to let users express the relative value of their traffic. Then, a connection with weight k may be equivalent to k connections of weight 1.

We illustrate the following definitions on the classical linear network represented in Figure 15.1. It consists of $|\mathcal{L}|$ links shared by $|\mathcal{L}|+1$ connections (numbered from 0 to $|\mathcal{L}|$) such that :

- each connection n, $n \neq 0$ goes only through link n,
- connection 0 uses all links.

Note that, for any Pareto optimal allocation, the global revenue can be written:

$$R_{glob} = \sum_{0 \leq n \leq |\mathcal{L}|} x_n = (1-|\mathcal{L}|)x_0 + \sum_{l \in \mathcal{L}} C_l. \tag{15.2}$$

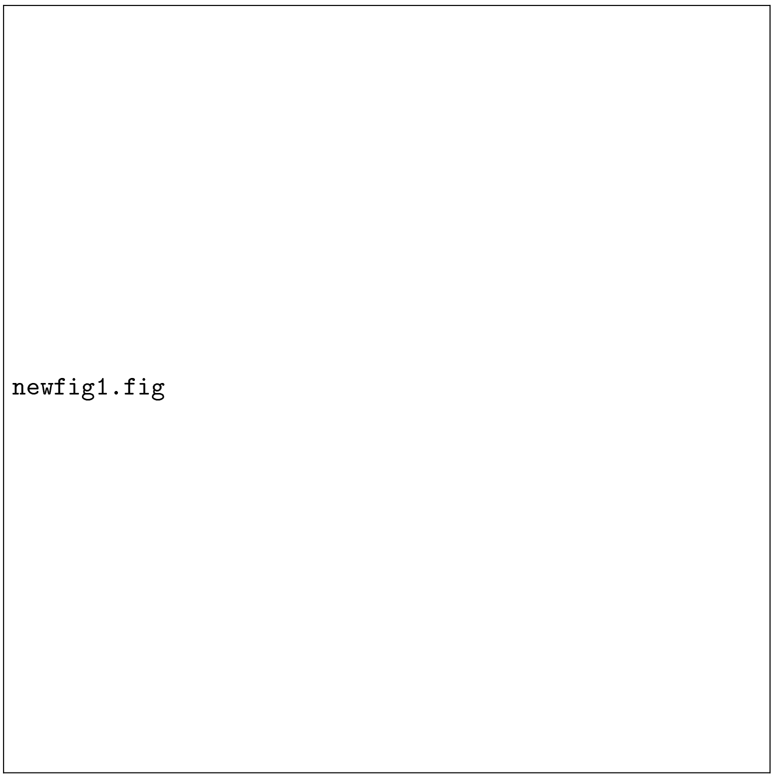


Figure 15.1: Linear network

15.3.1 Maximization of global throughput

This criterion consists of the maximization of the total allocated bandwidth $R_{glob} = \max \sum_{n \in \mathcal{N}} x_n$.

A major drawback of this criterion is that it can lead to situations in which the allocation is null for one or more connections, as we can see in the linear example (indeed, from Equation 15.2 the criterion imposes that $x_0 = 0$). This is why it is not considered as fair.

A variant, in which we assign to each user a utility function (we will see in Section 15.4.1 how this can be justified in the case of networks) has been developed. Roughly speaking, the utility function represents the satisfaction of a user from its allocation. This can actually be the perceived quality of an audio or video signal. Some algorithms ([?], [?]) were then developed to maximize the sum of the users' utilities, which is called *social-welfare* optimization.

An opposite approach is the max-min criterion that seeks to allocate the resources in the most homogeneous way, while being Pareto optimal.

15.3.2 Max-Min fairness

Several equivalent (as proved in [?]) definitions can be found. Let us first define the bottleneck of a connection.

Definition 15. *A link l is a bottleneck of connection n if:*

- *It is fully used,*
- *The allocation of connection n is the greatest among all connections using link l.*

We then have the following definitions:

Definition 16. *An allocation is max-min fair if and only if each source has a bottleneck.*

Definition 17. *The max-min allocation is maximal for the order $\preccurlyeq$.*

Definition 18. *The max-min fair allocation is such that any increase of the allocation for one connection would be at the expense of another connection whose allocation was already smaller. In other words, if $\vec{x}$ is the max-min allocation then $\forall \vec{y}, \vec{y} \neq \vec{x}$, $y_n > x_n \Rightarrow \exists s, x_s \leq x_n, y_s < x_s$.*

In the linear example, the allocation of connection 0 will be half of the smaller link capacity: $x_0 = C_{l_{min}}/2$. The total revenue is then: $R_{max-min} = (1 - \frac{|\mathcal{L}|}{2})C_{l_{min}} + \sum_{l \in \mathcal{L}} C_n$.

The max-min fairness has been adopted by the ATM forum for the Available Bit Rate (ABR) service of ATM ([?]). It has been argued that max-min fairness gives to much allocation to *long* connections (in our case the connection 0) and does not efficiently utilize bandwidth. Therefore, a new concept has been introduced to Kelly: the proportional fairness.

15.3.3 Proportional fairness

It has been defined by two equivalent ways (as proved in [?]).

Definition 19. *The proportional fair assignment maximizes $\max\limits_{\vec{x} \in X} \sum\limits_{n \in \mathcal{N}} ln(x_n)$.*

Definition 20. *Proportional fairness is the unique allocation $\vec{x}$ such that for any other allocation $\vec{x}'$, we have $\sum\limits_{n \in \mathcal{N}} \dfrac{x'_n - x_n}{x_n} \leq 0$.*

In the linear network, $x_0 = 1/\sum_{n \in \mathcal{N}} \frac{1}{C_n - x_0}$. Furthermore, if the links are identical with capacity C, then $x_0 = \dfrac{C}{|\mathcal{L}| + 1}$. We note that, as expected: $x_{0_{\text{global opt}}} < x_{0_{\text{prop fair}}} < x_{0_{\text{max-min}}}$. Moreover $R_{\text{global opt}} \geq R_{\text{prop fair}} \geq R_{\text{max-min}}$. The compromise between fairness and efficiently was the primary reason for the success of proportional fairness.

Its weighted version can be written. An allocation $\vec{x} \in X^{wpf}$ is weighted proportional fair if for any other allocation $\vec{x}^* \in X$ we have ([42, ?]):
$\sum_{n \in \mathcal{N}} w_n \frac{x_n^* - x_n}{x_n} \leq 0$. Equivalently: $x^{wpf} = \max\limits_{\vec{x} \in X} \sum\limits_{x \in \mathcal{N}} w_n \ln(x_n) = \max\limits_{\vec{x} \in X} \prod\limits_{n \in \mathcal{N}} x_n^{w_n}$.

The congestion control mechanism based on linear increased-multiplicative decreased window flow control can lead to proportionally fair allocations under certain conditions ([?]). Unfortunately this result does not hold for all versions of TCP, and in particular to TCP Reno. But it has been proved that TCP Vegas actually achieve proportional fairness [?].

15.3.4 Potential delay minimization

This criterion, introduced in [?], is defined by: $\min\limits_{\vec{x} \in X} \sum\limits_{n \in \mathcal{N}} \dfrac{1}{x_n}$. It is a minimization problem of an inverse function of the bandwidth allocated, that is to say the transfer time.

15.3.5 Relation between the different fairness criteria

Mo and Walrand ([?]) recently showed that these criteria can be written as a single optimization function. Let $\vec{x}_\alpha$ be the allocation solution of:

$$\vec{x}_\alpha = \max_{\vec{x} \in X} \sum_{n \in \mathcal{N}} \frac{{x_n}^\alpha}{1 - \alpha} \tag{15.3}$$

With $\alpha \geq 0$, $\alpha \neq 1$. Then, for specific values of α, $\vec{x}_\alpha$ corresponds to the different fairness criteria previously mentioned : the global maximization of the throughput ($\alpha = 0$), proportional fairness ($\alpha \to 1$), potential delay minimization ($\alpha = 2$) or max-min fairness ($\alpha \to \infty$)

Figure 15.3.5 represents the allocation to connection 0 when α grows from zero to infinity in the case of the linear network with equal link capacities. We clearly see the influence of the parameter α. As α increases, the differences between the allocation of the connections vanish, thus continuously deriving from a global optimization equilibrium to a max-min allocation.

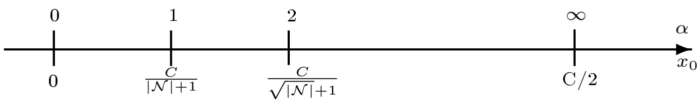


Figure 15.2: Allocation to connection 0 with α

15.4 Nash criterium

The Nash Bargaining Solution (NBS) [?] concept for fair allocation is frequently used in cooperative game theory. It is defined by a set of axioms that game theorists find natural to require in seeking fair allocations. These axioms deal with utilities associated to users, which is the natural interest of this approach compared to the previously seen fairness criteria.

The NBS has already been used (under a simplified version) in networks of general topologies to propose a fair share of resources ([?, ?]). But in both cases, the utility functions were linear. It has been shown that in fact, the utility functions can be any concave function.

Several fairness criteria were defined in cooperative game theory (see [?] and references herein). We consider here the Nash concept, since it can be seen, as previously mentioned, as a generalization of the widely studied proportional fairness.

Let us suppose that a finite number of perfectly rational individuals can collaborate in order to get mutual benefit. We further suppose that they can compare their satisfaction from the possession of the objects of the bargaining. We can then associate the users to a utility function, which is, obviously, not unique: if u is such a function, then $au + b$ is an equivalent one (for $a, b \in \mathbb{R}, a > 0$). In the case where the players cannot find an agreement, the game ends at the "disagreement point", characterized by a certain utility, u^0.

Let $X \subset \mathbb{R}^n$ denote the set of possible strategies. It is a convex closed and non empty set. The utility functions, $f_i : X \to \mathbb{R}, i = 1, \ldots, k$, are supposed to be upper bounded functions. The set of achievable utilities, U, $U \subset R^k$ such that $U = \{u \in \mathbb{R}^k | x \in X, u = (f_1(x), ..., f_k(x))\}$ is non empty, convex and closed and $u^0 \in \mathbb{R}^k$ is the utility from which the players accept to bargain. Finally, we denote by U^0 the set $U^0 = \{u \in U | u^0 \leq u\}$, the subset of U in which the players achieve more than their minimum requirements. Similarly, we define $X_0 = \{x \in X | \forall i, f_i(x) \geq u_i^0\}$.

Definition 21. *A mapping $S : (U, u^0) \to \mathbb{R}^n$ is said to be an NBP (Nash bargaining point) if:*

1. *it guarantees the* ***minimum required performances****: $S(U, u^0) \in U^0$.*
2. *$S(U, u^0)$ is* ***Pareto optimal****.*
3. *It is* ***linearly invariant****, i.e. the bargaining point is unchanged if the performance objectives are affinely scaled. More precisely, if $\phi : \mathbb{R}^n \to \mathbb{R}^n$ is a linear map such that $\forall i, \phi_i(v) = a_i v_i + b_i$, then $S(\phi(u), \phi(u^0)) = \phi(S(U, u^0))$.*
4. *S is* ***symmetric****, i.e. does not depend on the specific labels. Hence, connections with the same minimum performance u_i^0 and the same utilities will have the same performances.*
5. *S is* ***not affected by reducing the domain*** *if a solution to the problem with the larger domain can be found on the restricted one. More precisely, if $V \subset U$, and $S(U, u^0) \in V$ then $S(U, u^0) = S(V, u^0)$.*

Definition 22. *The point $u^* = S(U, u^0)$ is called the Nash Bargaining Point and $f^{-1}(u^*)$ is the set of Nash Bargaining Solutions.*

We have the equivalent optimization problem:

Theorem 20. *[?, Thm. 2.1, Thm 2.2] and [?]. Let the utility functions f_i be concave, upper-bounded, defined on X which is a convex and compact subset of $\mathbb{R}^n$. Let J be the set of users able to achieve a performance strictly superior to their initial performance, i.e. $J = \{j \in \{1, ..., N\} | \exists x \in X_0,\ s.t.\ f_j(x) > u_j^0\}$. Assume that $\{f_j\}_{j\in J}$ are injective. Then there exists a unique NBP as well as a unique NBS x that verifies $f_j(x) > u_j(x), j \in J$, and is the unique solution of the problem:*

$$\max \prod_{j\in J}(f_j(x) - u_j^0), \quad x \in X_0. \tag{15.4}$$

Equivalently, it is the unique solution of $\max \sum_{j\in J} \ln(f_j(x) - u_j^0), \quad x \in X_0.$

In 1991, [?] adapted the Nash bargaining solution to Jackson networks. Some years after, Yaïche et al. [?] adapted it for bandwidth allocation in networks. In their works however, they restricted themselves to linear utility functions.

We can note that the NBS corresponds to proportional fairness in the case where the utility functions are linear and where the MR_i are null.

Remark 10 (Other fairness criteria defined through a set of axioms)**.** *The last axiom (5) of the Nash Bargaining Point has suffered some criticisms, as it does not take into account how much the other players have given up. Two other interesting fairness criteria were then defined when modifying this last axiom, namely the Raiffa-Kalai-Smorondinsky and the Thomson (or "utilitarian choice rule") solution [?]. We do not treat here in more detail these two criteria as they correspond respectively of special cases of max-min fairness (Section 15.3.2) and the maximization of the global throughput (Section 15.3.1) when considering the utility of the applications.*

15.4.1 Utility functions in networking

The interest of the Nash bargaining concept, as opposed to the previous ones used so far in the telecommunication context is to take into account the different interests the users have for the shared resource[3]. In networks context, the bandwidth does not have the same value for different users. For instance, a user consulting his or her emails does not have the same needs that another one using phone over IP. The utility functions represent the impact of the bandwidth allocation on the perceived quality. We show in the following the shapes of the utility functions for different types of application. Our discussion is qualitative and inspired from the work of Shenker [?]. For numerical results, for instance on audio communications, the reader may refer to [?, ?]. We illustrate the different shapes of the utility functions on Figure 15.3.

Elastic applications have no real-time requirements and no rate constraints. Typical examples are file transfer or email. Their utility function is concave increasing without a minimum required rate.

"Delay adaptive" or "rate adaptive" applications have soft real-time requirements. Typical examples are voice or video over IP. In such applications, the compression rate of data is computed as a function of the quantity of available resource. The utility functions that we use to represent these applications are slightly different than those in [?]. In [?], the utility is strictly positive for any non zero bandwidth and tends to zero when the bandwidth does. We consider in contrast that the utility equals zero below a certain value, as in [?]. Indeed, in many voice applications, one can select the transmission rate by choosing an appropriate compression mechanism and existing compression software have an upper bound on the compression, which implies a lower bound on the transmission rate for which a communication can be initiated, which we denote *MR*. Thus, a maximum compression rate is associated with the lower acceptable quality for the user. If there is no sufficient bandwidth,

[3] Note, however, that in some cases are implemented *service classes*. They are so far limited to pricing purposes and the underlying idea is usually to give priority to some packets at the buffers.

the connection is not initiated. This kind of behavior generates utility functions that are null for bandwidth below MR and which are not differentiable at the point $(MR, 0)$. Similarly, it is useless to allocate a bandwidth greater than a certain threshold PR because the perceived gain for a human being will not be noticeable. As an example, for voice transmission, we usually consider throughputs in the range $[16, 40]$ kb/s. A user to whom we would allocate a throughput of 200 kb/s would not have a better quality feeling than that if its throughput was halved.

mfunctions3.fig

Figure 15.3: Utility function in networking

15.4.2 Proposed approximation scheme: quadratic utility functions

The utility functions of both “elastic traffic” and “delay adaptive” applications have a minimum value MR_n below which they equals zero (in the former case, $MR_n = 0$). As explained by Nash, the equilibrium point should not depend on the chosen representation of utility function. Therefore, we assume that $f_n(MR_n) = 0$. Beyond MR_n the function is concave and increasing with the bandwidth. We propose to approximate such a utility function with a parabola with parameters that depend on the applications (see Fig. 15.4). The general equation has the form:

$$f_n(x_n) = c_n - a_n(x_n - b_n)^2. \tag{15.5}$$

Note that the utility function is defined only until the point (PR_n, fPR_n), so we may ignore the whole right part of the parabola (and in particular, the part in which the function decreases).

We introduce:

- T_n the tangent of the utility function at the point $(MR_n, 0)$,
- fPR_n the utility value at point PR_n.

parabola.fig

Figure 15.4: Quadratic utility functions

f_n can be equally be defined by a_n, b_n, c_n or through the equations $f_n(MR_n) = 0$, $f_n(PR_n) = fPR_n$ and $f_n'(MR_n) = T_n$. We should note that, since PR_n is in the increasing part of the function, we have $\frac{1}{2}T_i(PR_i - MR_i) \leq fPR_i \leq T_i(PR_i - MR_i)$. We thus define the *concavity of the utility*, β_n, through

$$fPR_n = T_n \cdot \beta_n \cdot (PR_n - MR_n).$$

Note that $1/2 \leq \beta_n \leq 1$ and the smaller β_n is, the more concave is the utility. The limit $\beta_n = 1$ is the linear case (studied in [**?**]).

We can therefore equivalently use in the following the parameters a_n, b_n c_n or MR_i, T_i and β_i linked by the equations:

$$a_n = T_n \frac{1-\beta_n}{PR_n - MR_n},\ b_n = \frac{PR_n - (2\beta_n - 1)MR_n}{2(1-\beta_n)} \text{ and } c_n = \frac{T_n}{4}\frac{PR_n - MR_n}{1-\beta_n}.$$

15.5 Proposed fairness scheme

As previously mentioned, we propose to apply the classic fairness criteria used in networks to the utility functions of applications rather than to the throughput. We saw that, by applying this philosophy to the proportional fairness criterium, we obtain the Nash bargaining solution, historically defined by a set of axioms rather than an optimization problem.

We have presented a family of fairness criteria, parameterized by a real α ($\alpha > 0, a \neq 1$) that allows an network operator to choose an equilibrium between fairness and resource utilization.

When applied to the utilities of the connections, the optimization problem becomes:

$$\max_{n \in \mathcal{N}} \frac{[f_n(x_n)]^{1-\alpha}}{1-\alpha}, \text{if } a > 0, a \neq 1 \text{ and } \prod f_n(x_n) \text{ if } \alpha = 1 \tag{15.6}$$

subject to the problem constraints.

We also propose to approximate utility functions by quadratic parameterized functions. Note that, as in our approximation $f(MR) = 0$, the proposed criterion coincides with the NBS when $\alpha \to 1$. The advantage of using quadratic utilities is that it allows one to approximate the concave real utilities drawn from experiments with functions that are simple to handle through a small set of parameters. Also, it allows us to easily analyze the influence of these parameters on the obtained equilibrium. Of particular interest is the influence of the concavity.

15.5.1 Property: influence of the concavity

We study the impact of concavity on the NBS. Consider two differentiable functions f and g defined on the same interval $(MR, PR]$ where both are strictly positive on $]MR, PR]$.

Definition 23. *We say that f is more concave than g if for every $x \in (MR, PR]$ the relative derivative of f is smaller than or equal to that of g, i.e. $f'(x)/f(x) \le g'(x)/g(x)$. If f or g are not differentiable at x, one could require instead that the same relation holds for the supergradients: if $\hat{f}(x)$ is the largest supergradient of f at x and $\hat{g}(x)$ is the smallest supergradient of g at x, then we require $\hat{f}(x)/f(x) \le \hat{g}(x)/g(x)$.*

Motivated by (15.4), we say that :

Definition 24. *An assignment $\vec{x}$ is more fair in the sense of NBS than an assignment $\vec{y}$ if $\prod_{n\in\mathcal{N}}(f_n(x_n) - f_n(MR_n)) \ge \prod_{n\in\mathcal{N}}(f_n(y_n) - f_n(MR_n))$.*

Consider the case in which $f_n(MR_n) = 0$. Consider 2 connections with utilities f and g as above competing for the bandwidth of a single link of capacity C. If we had ignored the utilities of the connections, we would have assigned them an equal bandwidth (according to the original proportional allocation), which we denote by $x = C/2$.

Proposition 1. *If two connections are competing for the bandwidth of a single link, the fair assignment has the property that more bandwidth is assigned to the less concave function.*

Proof: We show that by transferring bandwidth from the connection with the more concave utility (say f) to the other one, we improve the fairness (assuming this does not violate the MR and PR constraints) in the sense of the NBS. Indeed, we have

$$g(x+\epsilon)f(x-\epsilon) = g(x)f(x)\left(1 + \epsilon\left[\frac{g'(x)}{g(x)} - \frac{f'(x)}{f(x)}\right] + o(\epsilon)\right).$$

We conclude that there is some ϵ_0 s.t. $\forall\epsilon < \epsilon_0$, $g(x+\epsilon)f(x-\epsilon) > g(x)f(x)$. Hence we strictly improve the fairness by transferring an amount of ϵ_0 to the connection with less concave utility. By further increasing this amount, we shall eventually reach a local maximum (since our function is continuous over a compact interval). This will be a global maximum since (15.4) is a maximization problem of a concave function over a convex set. ∎

Example Let two connections of bandwidth x_1 and x_2 sharing a single link of capacity C. Their utility functions f and g are represented in Figure 15.5.
For $x \ge 0$, $f(x) = 3x\mathbb{1}_{0\le x\le 1} + (2+x)\mathbb{1}_{x>1}$, and $g(x) = 2x$.
Then $f'(x)/f(x) = \begin{cases} x^{-1} & \text{for } x \in [0,1[, \\ (2+x)^{-1} & \text{for } x \ge 1, \end{cases}$ whereas $\forall x, g'(x)/g(x) = x^{-1}$.
(At $x = 1$, f is not differentiable but its supergradients at that point constitute the set $[1/3, 1]$).
Thus f is more concave than g. We assume that $PR_1 + PR_2 > C$. The NBS is the argument of

$$\zeta(C) = \max f(x)g(C-x) = \max\left(\max_{x\in[0,1]} h(x), \max_{x>1} k(x)\right)$$

$$\text{with } h(x) = 6x(C-x), \text{ and } k(x) = 2(x+2)(C-x).$$

Proposition 2. *The NBS is depicted in Fig 15.6. We distinguish 3 regions:*
(i) $C < 2$, where $\zeta(C) = 3c^2/2$ and the NBS is $(x_1^*, x_2^*) = (C/2, C/2)$,
(ii) $2 \le C < 4$, where $\zeta(C) = 6(C-1)$ and $(x_1^*, x_2^*) = (1, C-1)$,
(iii) $C \ge 4$, where $\zeta(C) = 2(c/2+1)^2$ and $(x_1^*, x_2^*) = (C/2-1, C/2+1)$.

We see in this example that the least concave function receives at least as much as the other one, and the difference increases with C. It is impressive to note that there is a region in which an increase in the capacity benefits only to one connection. The example illustrates the power

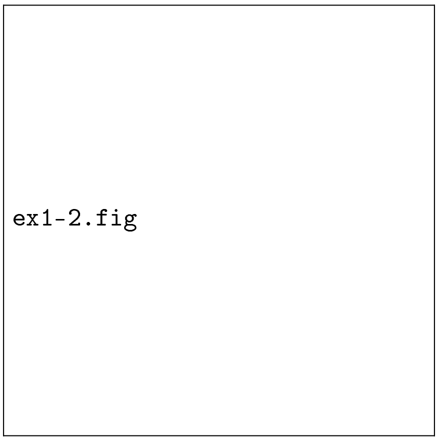


Figure 15.5: Utility functions.

Figure 15.6: NBS of two connections sharing a link.

of the NBS approach: the original proportional fairness, or even weighted proportional fairness, would assign a proportion of the capacity to each connection that does not vary as we increase the capacity, since it is insensitive to the utilities. In contrast, utility sensitive fairness concepts allocate the bandwidth in a dynamic way: the proportion assigned to each connection is a function of the capacity.

For the experimental part, we propose to study 2 different fairness scenarios, that we present in the rest of this section.

15.5.2 Possible optimizations

We propose two types of optimizations, called *connection-aware* and *network-aware*. The idea in the first is to optimize the allocation to the connections and in the former to better utilize the network resources.

Connection-aware optimization We associate to each connection a utility function and, once the minimal requirements are satisfied, fairly allocate the extra bandwidth among the users. The $\vec{\kappa}$ vector is given and the optimization problem is $\max \sum_{n \in \mathcal{N}} \frac{f_n(x_n)^{1-\alpha}}{1-\alpha}$ with the system's constraints previously mentioned.

Network-aware optimization We apply the fairness concepts to the remaining bandwidth available in the links. Indeed, it can be interesting to maximize the bandwidth available in a link in order to use it for instance for some other traffic type (this is of particular interest if the routing policies for connections cannot be dynamically modified).

Therefore, $\vec{x}$, the bandwidth allocation to connections, is given. We then consider the fair allocation of remaining bandwidth, that is to say of vector $\vec{\kappa}$. The problem can be written $\max \sum_{l \in \mathcal{L}} \frac{f_l(\kappa_l)^{1-\alpha}}{1-\alpha}$. We can note that the allocation vector being given, the optimization problem is of interest only is we consider the fractional-routing. Indeed, if the routing is fixed, then the quantity available in each link is fixed as well.

We make the following assumption in the rest of the article: The network has sufficient capacity to satisfy all the minimum requirements.

In the next section we propose an example of explicit rate computation. This analytic study is possible, however, only in the case of simple network topologies and rely on specific assumptions.

15.6 Explicit computation of rates

In this section, we study the fair allocations in the linear network and its generalization, the grid network.

15.6.1 The linear network

We again consider the linear network. We use the obvious associated routing and consider in the following the connection-aware optimization with fixed routing constraints.

We recall that the fairness equilibria is obtained through the optimization of: $\frac{1}{1-\alpha}\sum_{n\in\mathcal{N}}(f_n(x_n))^{1-\alpha}$ if $\alpha \neq 1$, and $\prod_{n\in\mathcal{N}} f_n(x_n)$otherwise with the constraints:

$$\begin{cases} \forall n \in \mathcal{N}, MR_n \leq x_n \leq PR_n & \text{(user's constraint)}, \\ \forall n \in \mathcal{L}, x_0 + x_n \leq C_n & \text{(capacity constraints)}. \end{cases} \tag{15.7}$$

Note that in this network all the links can be saturated: $\forall n \in \mathcal{L}, x_0 + x_n = C_n$.

We make two significant assumptions. First, that each link has the same capacity C. Secondly, we suppose that each of the connections $n, n \neq 0$ has the same utility function: $\forall n \in \mathcal{L}, a_n = a_1, b_n = b_1, c_n = c_1$. Thus the fair allocation will assign the same bandwidth to each connection $n, n \neq 0$.

Therefore, by denoting by x the bandwidth allocated to connection 0, the fair allocation problem becomes the maximization of (15.8) under the constraints (15.7).

$$\begin{cases} g(x) = \dfrac{1}{1-\alpha}\left[(f_0(x))^{1-\alpha} + |\mathcal{L}|(f_1(C-x))^{1-\alpha}\right] & \text{if } \alpha \neq 1 \\ h(x) = f_0(x)(f_1(C-x))^{|\mathcal{L}|} & \text{otherwise.} \end{cases} \tag{15.8}$$

Remark 11. *The problem is of interest only if:*

$$\begin{cases} MR_0 + MR_1 \leq C & \textit{(feasible system, Hypothesis 15.5.2)}, \\ PR_0 + PR_1 > C & \textit{(not all maximum demands can be satisfied)}. \end{cases}$$

Solution of the linear problem. By differentiating (15.8) we obtain:

$$\begin{cases} a_0(x-b_0)f_1(C-x)^{\alpha} = |\mathcal{L}|a_1(C-x-b_1)f_0(x)^{\alpha} & \text{if } \alpha \neq 1, \\ a_0(x-b_0)f_1(C-x)^{|\mathcal{L}|} = |\mathcal{L}|a_1(C-x-b_1)f_1(C-x)^{|\mathcal{L}|-1}f_0(x) & \text{otherwise.} \end{cases}$$

We note that for $\alpha = 1$, $f_1(C-x) = 0 \Rightarrow h(x) = 0$. Therefore $f_1(C-x) \neq 0$ and the bandwidth x associated to connection 0 satisfies, for all α:

$$a_0(x-b_0)(f_1(C-x))^{\alpha} = |\mathcal{L}|a_1(C-x-b_1)(f_0(x))^{\alpha}. \tag{15.9}$$

Limits and asymptotic analysis

Proposition 3. *When $|\mathcal{L}|$ grows to infinity, the allocation of connection* 0 *tend to* $\max(MR_0, C-b_1)$.

The proof is given in Appendix. Note that the limit does not depend on the concavity of the utility of connection 0. We show in Figure 15.6.1 how the system converge to x_{lim} as $\mathcal{L}$ grows to infinity in the case of the NBS ($\alpha = 1$) and $C - b_1 \geq MR_0$.

We further refine the analysis of the limit when $|\mathcal{L}|$ becomes large:

Proposition 4. *If $C - b_1 < MR_0$ then x is such that*

$$x - MR_0 \sim Z \textit{ with } \quad Z^\alpha \sim_{l\in\mathcal{L}} \frac{1}{|\mathcal{L}|}\left(\frac{1}{2(b_0 - MR_0)\cdot a_0}\right)^{\alpha-1}\frac{f_1(C-MR_0)^\alpha}{f_1'(C-MR_0)}.$$

$$\textit{Otherwise, } x = C - b_1 + Z + o(1/|\mathcal{L}|) \textit{ with } \quad Z = \frac{1}{|\mathcal{L}|}\frac{c_1^\alpha}{2a_1}\frac{f_0'(C-b_1)}{f_0(C-b_1)^\alpha}.$$

Proof: If $C - b_1 < MR_0$, then, as $|\mathcal{L}| \to \infty$, the left hand side of (15.9) tends to a non-null constant: $\lim\limits_{|\mathcal{L}|\to\infty} a_0(x-b_0)(f_1(C-x))^\alpha = a_0(MR_0 - b_0)(f_1(C-MR_0)^\alpha)$.
We now examine the right hand side of (15.9). It can be written as:

$$\begin{aligned}|\mathcal{L}|a_1(C-x-b_1)f_0(x)^\alpha &= |\mathcal{L}|a_1(C-b_1-MR_0-z)(c_0 - a_0(MR_0+z-b_0)^2)^\alpha\\ &= |\mathcal{L}|a_1(C-b_1-MR_0-z)(2a_0(b_0-MR_0)z - a_0z^2)^\alpha\\ &\sim_{|\mathcal{L}|\to\infty} |\mathcal{L}|a_0^\alpha a_1 2^\alpha(C-b_1-MR_0)(b_0-MR_0)^\alpha z^\alpha.\end{aligned}$$

which yields (4) by substituting the appropriate expressions.

If $MR_0 \leq C - b_1$, as $|\mathcal{L}| \to \infty$, the left hand side tends to the constant $a_0(C-b_1-b_0)c_1^\alpha$. The right hand side is: $-|\mathcal{L}|a_1 z\Big(c_0 - a_0(C-b_1-b_0+z)^2\Big)$. Hence the result. ∎

Remark 12. *In (4), we can check the asymptotes for special cases (and for $\alpha = 1$):*

- *If $\beta_1 \to 1$ then $Z = \dfrac{C - MR_0 - MR_1}{|\mathcal{L}|}$ (linear case, already obtained in [?]).*
- *If $\beta_1 \to 1/2$ then $Z = \dfrac{C - MR_0 - MR_1}{2|\mathcal{L}|}\left[1 - \dfrac{PR_1 - MR_1}{C - MR_0 - PR_1}\right]$.*

We studied the case of a linear network with two kinds of connections. In the case where $\alpha = 1$ (NBS), we note (from (15.9)) that the problem is a third order polynomial and can therefore be explicitly solved. If $\alpha = 0$ (global optimization), the allocation is given by the solution of a first order polynomial. In any other case, only numerical results can be used. Still, we can make the limits of the allocations explicit for any value of α, and give an equivalent when the number of links becomes large.

15.6.2 Grid network

This network is the natural generalization of the linear network. It consists of $K \times L$ capacity links with K horizontal routes and L vertical routes as shown in Figure 15.7. We focus here on the fixed routing policy.

As in the previous example, we also suppose that, for each $i \in \{1,\dots L\}, j \in \{1,\dots K\}$, $MR_i + PR_{L+j} \geq C_{i,L+j}$ and $PR_i + MR_{L+j} \geq C_{i,L+j}$. We further suppose that the links have equal capacity. We suppose that all the horizontal connections (respectively vertical) have the same utility function f_h (respectively f_v). We can then conclude that all the horizontal connections (vertical) will get the same throughput x ($x_v = C - x$). We then wish to maximize, in the case of the NBS:

$$\prod_{n\in[1:L]\bigcup[L+1,L+K]} f_n(x_n) = (f_h(x))^K * (f_v(C-x))^L. \tag{15.10}$$

Proposition 5. *In the grid network, if $C - b_1 < MR_0$, x verifies: $x - MR_h \sim Z$ with: $Z = \frac{K}{L}\left(\frac{1}{2(b_h - MR_h)a_h}\right)^{\alpha-1}\frac{f_v(c-MR_0)}{f_v'(c-MR_0)}$. Otherwise, $x = C - b_1 + Z + o(K/L)$ with $Z = \frac{K}{L}\frac{c_v^\alpha}{2a_v}\frac{f_h'(c-b_v)}{f_h'(c-b_v)}$. Note that if $L = K$ and $f_h = f_v$ we obtain $x = C/2$.*

Proof: Maximizing Equation 15.10 is similar to maximizing $f_h(x) * (f_v(C-x))^{L/K}$. This is equivalent to the linear problem by substituting $|\mathcal{L}|$ to L/K. ∎

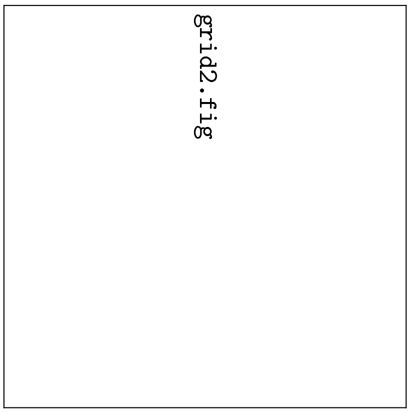


Figure 15.7: A grid network.

15.7 Lagrangian method

The Lagrangian method was proposed by [?] to obtain the NBS for the special case of linear utility function. It has the advantage of having distributed implementations. We generalize below this approach to the quadratic utility, for which the linear case can be recovered by taking $\beta \to 1$.

Unfortunately, the resulting decentralized iterative algorithm of the problem proposed in [?] is not fully satisfactory for quadratic utilities, as we will see in the following.

15.7.1 Lagrangian multipliers

We now use the Kuhn-Tucker conditions for (15.6) to obtain an alternative characterization of the NBS in terms of the corresponding Lagrange multipliers under the hypothesis of fixed routing and the connection aware optimization.

Proposition 6. *Under the hypothesis that $\forall l \in \mathcal{L}, \sum a_{ln} MR_n < C_l$ (this amounts in saying that no link is saturated[4]), the NBS is characterized by: $\exists \mu_l \geq 0, l \in \mathcal{L}$ such that $\forall n \in \mathcal{N}$, we have*

$$x_n = \min\left(PR_n, MR_n + \left(\sum_{l\in\mathcal{L}} \mu_l a_{l,n}\right)^{-1} + \frac{1}{2} \cdot \frac{PR_n - MR_n}{1-\beta_n} \times \left[1 - \sqrt{1 + \frac{4\left(\frac{1-\beta_n}{PR_n - MR_n}\right)^2}{\left(\sum_{l\in\mathcal{L}} \mu_l a_{l,n}\right)^2}}\right]\right).$$

Proof: Under the assumption $\sum_{n\in\mathcal{N}} a_{ln} MR_n < C_l$, the set of possible solutions of (15.6) is non-empty, convex and compact. The constraints (as defined in Subsection 15.2.3) are linear in x_n and $f(x) = \sum_{n\in\mathcal{N}} \ln f_n(x_n)$ is C^1. Therefore the first order Kuhn-Tucker conditions are necessary and sufficient for optimality. The Lagrangian associated with (15.6) is

$$\mathcal{L}(x,\lambda,\delta,\mu) = f(x) - \sum_{n\in\mathcal{N}} \lambda_n (MR_n - x_n) - \sum_{n\in\mathcal{N}} \delta_n (x_n - PR_n) - \sum_{l\in\mathcal{L}} \mu_l ((Ax)_l - C_l).$$

[4] This assumption does not constraint our problem. If a link is saturated when considering the minimal demands of the connections, then we rewrite the system by suppressing this link and the connections that were going through it. The bandwidth that these connections were using is then a new capacity constraint on the other links of the network.

For $n \in \mathcal{N}$, $\lambda_n \geq 0$ and $\delta_n \geq 0$ are the Lagrange multipliers associated with the constraints $x_n \geq MR_n$ and $x_n \leq PR_n$ respectively. $\mu_l \geq 0, l \in \mathcal{L}$ are the Lagrange multipliers associated with the capacity constraints. The first order optimality conditions are thus: $\forall n \in \mathcal{N}$,

$$\frac{f_n'(x_n)}{f_n(x_n)} = \delta_n + \sum_{l \in \mathcal{L}} \mu_l A_{l,n} - \lambda_n, \quad (x_n - MR_n)\lambda_n = 0, \quad (x_n - PR_n)\delta_n = 0,$$

$$\text{and } \forall l, l \in \mathcal{L}, ((A\vec{x})_l - C_l)\mu_l = 0$$

Moreover, $\sum a_{ln} MR_n < C_l$ implies that $\forall n, \lambda_n = 0$ as in [?], and either $x_n = PR_n$ or $\delta_n = 0$, which yields the conclusion. ∎

Remark 13. *As $\beta \to 1$ we obtain the solution of [?] corresponding to linear utility:* $x_n = \min\left(PR_n, MR_n + \left[\sum_{l \in \mathcal{L}} \mu_l a_{l,n}\right]^{-1}\right)$.

$\mu_l, l \in \mathcal{L}$ represent the implied cost associated with the network link l. They represents the marginal cost of a rate unit allocated for any connection crossing link l.

15.7.2 Dual problem

Once we have explicitly expressed the NBS in terms of the Lagrange multipliers, we can actually solve the NBS using the dual problem in which we compute the Lagrange multipliers. The dual problem is :

$$\max_{\mu \in \mathbb{R}_+^{\mathcal{L}}} d(\mu) \quad \text{with} \quad d(\mu) = \min_{\vec{x} \in X} L(\vec{x}, \mu) = L(\overline{x}, \mu)$$

if we denote by $\overline{x}$ the optimal allocation. The vector $\overline{x} = \overline{x_1, x_2 ... x_n}$ is the NBS. We now use the result obtained in the primal, and, for a given vector μ, we note for each connection n: $x_n(\mu) = g_n(\sum_{l \in \mathcal{L}} \mu_l \cdot a_{l,n})$.

$$\text{With } g_n(p) = \begin{cases} PR_n & \text{if } p \leq \dfrac{2\beta_n - 1}{\beta_n} \dfrac{1}{PR_n - MR_n}, \\ MR_n + \dfrac{1}{p} + \dfrac{PR_n - MR_n}{2(1-\beta_n)} \left[1 - \sqrt{1 + \left(\frac{2}{p} \frac{1-\beta_n}{PR_n - MR_n}\right)^2}\right] & \text{otherwise.} \end{cases}$$

We obtain for each $\mu \in \mathbb{R}^{\mathcal{L}}$:

$$d(\mu) = \sum_{n \in \mathcal{N}} -\ln(f_n(g_n(\sum_{l \in \mathcal{L}} \mu_l a_{l,n}))) + \sum_{l \in \mathcal{L}} \mu_l \sum_{n \in \mathcal{N}} a_{l,n} g_n\Big(\sum_{l \in \mathcal{L}} \mu_l a_{l,n}\Big) - \sum_{l \in \mathcal{L}} C_l \mu_l.$$

The idea is then to use a appropriate step γ and choose at each iteration:

$$\mu_l^{(k+1)} = \max\left(0, \mu_l^{(k)} + \gamma \frac{\partial d}{\partial \mu_l}\right). \tag{15.11}$$

But, if the utilities are not linear, then the partial derivative of d depends on their parameters. Therefore, the dual cannot be used to obtain a fully distributed algorithm anymore since every link needs to know the utility functions and allocations of all the connections of the system.

The Lagrangian relaxation let us introduce a optimization problem where each user computers its allocation x_n

$$x_n = g_n(\sum \mu_l a_{l,n})$$

while the network computes the link prices μ_l iteratively from Equation 15.11.

In the case of proportional fairness ($\alpha = 1$) with linear utility functions, the dual formulation can be used to obtain a decentralized algorithm. Unfortunately, the use of quadratic utilities require the knowledge in the network of all the connections and parameters of connections, which makes this method not suitable for large systems.

15.8 A Semi-Definite Programming (SDP) solution

In this section, we propose an alternative centralized method for solving the general fairness problem (15.6). It uses a mathematical program called Semi-Definite Program, which can be solved in polynomial time in theory and is tractable in practice. One can then use public domain programs to solve SDP[5]. SDP solves the minimization problem of a linear combination of variables (given by the scalar product of a vector L and the vector of variables) subject to a constraint of positive semi-definiteness (psd) of some general symmetric matrix P whose entries are either variables or constants. More details on SDP can be found in [?]. We study the complexity of our method and show that the proposed construction of the SDP matrix allows us to analyze large networks.

In the first subsection, we see how SDP was defined as an extension from linear programming and then recall some basic results of linear algebra. We describe the general shape of our SDP matrices to solve our optimization problem (Subsection 15.8.2), and the details for the different values of α (Subsection 15.8.3). We finally present a detailed description for a simple example (Subsection 15.8.4).

15.8.1 From linear programming to SDP

Definition 25 (Linear programming). *A linear program (LP) has the form:*

$$\min\{c^T x | Ax \geq b\}$$

where x is the unknown vector, c is a given coefficient vector, $A \in M_{m,n}$ is a $m \times n$ matrix of constraints and b is the vector of constraints.

We can extend linear to conic programming. In any Euclidian Space E we can define convex pointed cones ([?], [?]). Then, each cone $\mathcal{K}$ induce a partial order in E, that we denote $\geq_{\mathcal{K}}$: $a \geq_K b \Leftrightarrow a - b \geq_K 0 \Leftrightarrow a - b \in \mathcal{K}$.

Definition 26 (Conic programming). *Let $\mathcal{K}$ be a convex pointed cone of E, $c \in \mathbb{R}^n$ an objective vector, $b \in E$ and A a linear mapping $A : \left\{ \begin{array}{l} x \mapsto Ax \\ \mathbb{R}^n \to E. \end{array} \right.$. Then, the optimization problem* $\min_x \{c^T x | Ax \geq_K b\}$ *is a conic problem.*

Semi Definite Programming refers to the problems associated to $\mathcal{K} = S_m^+$ the cone of semi-definite positive matrices in the Euclidian space $E = S^m$ the set of symmetric matrices of size $m \times m$.

Definition 27. *A symmetric matrix A of size $m \times m$ is said semi-definite positive and we note $A \succcurlyeq 0$ if $\forall x \in \mathbb{R}_m, x^T Ax \geq 0$. Equivalently, all its eigenvalues[6] are positive or null.*

Definition 28 (Semi-Definite Programming). *A semi-definite programming problem is an optimization problem of the form*

$$\min c^T.vec(X) \quad \text{such that} \left\{ \begin{array}{l} A.vec(X) = b \\ X \succcurlyeq 0. \end{array} \right.$$

with $X \in S_n^+$, $c \in \mathbb{R}^{n^2}$, $b \in \mathbb{R}^m$ and $A \in M_{m,n^2}$. [7]

The following are results of linear algebra needed to understand our solving method. The proofs can be found in [?].

Proposition 7 (Real symmetrical matrices.). *All the eigenvalues of a real symmetric matrix are real.*

[5] see http://www.cs.nyu.edu/cs/faculty/overton/sdppack/sdppack.html
[6] The eigenvalues are the solutions of the polynomial $\det(A - \lambda I)$.
[7] If X is the matrix $X = (x_{i,j}), 1 \leq i,j \leq n$ then $vec(X)$ is the vector $(x_{1,1}, \ldots, x_{1,n}, x_{2,n}, \ldots x_{n,n})$.

Proposition 8 (Symmetric matrices of size 2)**.** *Let a, b and c be real positive numbers. Then, the matrix $M = \begin{pmatrix} m_{1,1} & m_{1,2} \\ m_{2,1} & m_{2,2} \end{pmatrix}$ is positive semi-definite (psd) if and only if $m_{1,1}m_{2,2} \geq M_{1,2}^2$.*

Proposition 9 (Symmetric matrices of size 1)**.** *A matrix of size 1 (a scalar) is positive if and only if its unique element is positive or null. $M = (m_{1,1}) \succcurlyeq 0 \Leftrightarrow m_{1,1} \geq 0$.*

Proposition 10. *A bloc diagonal matrix[8] is psd if and only if all its blocs are psd.*

15.8.2 General solving method with SDP

The general idea of SDP is to transform the original maximization of a *function* into a minimization problem of some new *variable* (or more generally of a linear combination of variables) subject to a constraint of positive semi-definiteness (psd) of some general matrix P. The positiveness (psd) of the matrix will:

- insure the system constraints,
- replace the objective function (of (15.6)) by a single variable.

We will see that in our case, we can construct the matrix with blocs of size 1 and 2. Note that the order of the blocs in the matrix is not relevant. Note further that the number of variables cannot be known a priori. Indeed, the creation of the matrix and of the blocs induce the creation of several intermediate variables that we will precise in the following.

Finally, the entries we will give to the SDP solver are the matrix and a vector of variable. In our case, this vector will be composed of "0" at each but one entry, corresponding to the final variable we want to maximize (this entry will contain the value "-1").

In the following two paragraphs, we explicit the blocs expressing the system constraints and those that enable us to transform the users utility functions into variables.

System constraints As seen in Subsection 15.2.3 the system constraints are linear. From Proposition 9 these constraints can be expressed via matrices of size 1 (scalar matrices).

We give in this paragraph the constraints representing the system with fixed routing (the fractional routing is similar). The capacity constraints are given by the positivity of $|\mathcal{L}|$ matrices, corresponding to the $|\mathcal{L}|$ links. Indeed, we recall that these constraints are $\forall l \in \mathcal{L}, (Ax)_l \leq C_l$. Then, it is sufficient to introduce $|\mathcal{L}|$ scalar matrices $capa^l$, such that: $\forall l \in \mathcal{L}, capa^l_{1,1} = C_l - (Ax)_l$. The users constraints are given by: $\forall n \in \mathcal{N}, MR_n \leq x_n \leq PR_n$. We therefore introduce $2 \times |\mathcal{N}|$ scalar matrices $util$ so that $util^i_{1,1} = PR_i - x_i$ if $i \leq |\mathcal{N}|$ and $util = x_i - MR_i$ if $|\mathcal{N}| < i \leq 2|\mathcal{N}|$.

Relations between the variables The initial variables are the bandwidth allocated to connections $x_n, n \in \mathcal{N}$. We introduce the intermediate variables $w_n, n \in \mathcal{N}$ with the following proposition:

Proposition 11. *Let x_n be a positive number and f_n its utility function, as defined in Equation 15.5. We introduce $F_n = \begin{pmatrix} -\dfrac{w_n - c_n}{a_n} & x_n - b_n \\ x_n - b_n & 1 \end{pmatrix}$. Then $F_n \succcurlyeq 0 \Leftrightarrow w_n \leq f_n(x_n)$.*

By using the blocs $F_n, n \in \mathcal{N}$, we replace the initial variables $x_n, n \in \mathcal{N}$ by $w_n, n \in \mathcal{N}$ and instead of maximizing a function of x_n, we maximize a function of w_n. Indeed, by distinguishing three cases depending of the sign of $\alpha - 1$, we can prove that $\forall \alpha \geq 0, \frac{1}{1-\alpha} \sum w_n^{1-\alpha} \leq \frac{1}{1-\alpha} \sum f_n(x_n)^{1-\alpha}$.

[8] A bloc diagonal matrix has the form $A = \begin{pmatrix} \boxed{A_1} & & & 0 \\ & \boxed{A_2} & & \\ & & \ddots & \\ 0 & & & \boxed{A_n} \end{pmatrix}$.

We can note, from this simple example that we replaced the utility functions by a constraint of psd of a matrix. Let us suppose we have $\mathcal{N}$ variables y_n such that $\forall n \in \mathcal{N}, y_n \leq \frac{w_n^{1-\alpha}}{1-\alpha}$ (for a given α). Then, the psd of scalar matrix END, defined by $END_{1,1} = (\sum_{k\in\mathcal{N}} y_k - z)$ insures $z \leq \sum \frac{w_n^{1-\alpha}}{1-\alpha}$ and maximizing z leads to the desired optimization since $z \leq \sum_n y_n \leq \frac{w_n^{1-\alpha}}{1-\alpha} \leq \frac{f_n(x_n)^{1-\alpha}}{1-\alpha}$.

We provide, in the following subsection a method to construct matrices whose psd constraint will insure that (we omit the "n" subscript):

$$y \leq w^{1-\alpha} \quad \text{if } 1-\alpha > 0, \qquad y \geq w^{1-\alpha} \quad \text{otherwise.}$$

15.8.3 Different values of α

The first paragraph deals with the case $\alpha \neq 1$ whereas second concerns the most complex case of the NBS ($\alpha = 1$).

Case $\alpha \neq 1$

Remark 14 (Case $\alpha = 2$). *The positiveness of matrix* $H = \begin{pmatrix} w & 1 \\ 1 & y \end{pmatrix}$ *ensures that* $y \geq 1/w$.

We use an idea of Nemirovski to provide a resolution method with a good approximation for any value of $\alpha > 0$ with $\alpha \neq \{1, 2\}$.

If $0 < \alpha < 1$, we have $1 - \alpha > 0$. Therefore, it is sufficient to provide one or several matrices whose psd will ensure that $y \leq w^{1-\alpha}$.

Proposition 12 (Case $0 < \alpha < 1$). *Then* $1 - \alpha < 1$ *and*

$$\forall \varepsilon > 0, \exists p \in \mathbb{N}, k \in \{0, \ldots, 2^p - 1\}, |(1-\alpha) - k/2^p| \leq \varepsilon.$$

Let $w, y \in \mathbb{R}^+$. *It is possible, using SDP constraints, to bound* y *and* w *by the relation* $y \leq w^{k/2^p}$.

Proof: Let $c_1, \ldots, c_p$ a serie of $0/1$ integers, such that $k = \sum_{i=1}^{p} c_i 2^{i-1}$. We note $y_0 = 1$ and submit $y_1, \ldots, y_p$ to the following constraints:

$$\begin{pmatrix} y_{i-1} & y_i \\ y_i & w \end{pmatrix} \text{if } c_i = 1 \text{ and } \begin{pmatrix} y_{i-1} & y_i \\ y_i & 1 \end{pmatrix} \text{if } c_i = 0.$$

Then obviously $y_i^2 \leq y_{i-1} w^{c_i}$, and $y_p \leq w^{k/2^p}$. Hence the result, by setting $y_p = y$. ∎

If $\alpha > 1$ we aim at finding matrices those psd constraint will ensure that $y \geq w^{1-\alpha}$.

Proposition 13 (Case $1 < \alpha < 2$). *Since* $0 < -(1-\alpha) < 1$ *we have:*

$$\forall \varepsilon > 0, \exists p \in \mathbb{N}, k \in \{0, \ldots, 2^p - 1\}, |-(1-\alpha) - k/2^p| \leq \varepsilon.$$

Let $w, y \in \mathbb{R}^+$. *It is possible, using SDP constraints, to bound* w *and* y *by the relation* $w \geq y^{-k/2^p}$.

Proof: Let c be an intermediate variable. Using Proposition 12, one can set $c \leq y^\beta$. Also, one can write: $\begin{pmatrix} y & 1 \\ 1 & w \end{pmatrix} \succcurlyeq 0$ which leads to $yw \geq 1$. Then $yw^\beta \geq 1$. ∎

The following proposition covers the cases $\alpha \in [2; +\infty[$.

Proposition 14 (Case $2 < \alpha$). *We have* $0 < \frac{1}{1-\alpha} < 1$. *Then*

$$\forall \varepsilon > 0, \exists p \in \mathbb{N}, k \in \{0, \ldots, 2^p - 1\}, \left|\frac{1}{1-\alpha} - k/2^p\right| \leq \varepsilon.$$

Let $y, w \in \mathbb{R}^+$. *It is possible, using SDP constraints to bound* w *et* y *by the relation* $y \geq w^{2^p/k}$.

Proof: Similarly to the proof of Proposition 12, we obtain $wy^\beta \geq 1$. ∎

Computing the NBS (case $\alpha = 1$)

We propose in this section a method to compute the NBS (case $\alpha = 1$). The method used is slightly different than from the other cases since we don't use the intermediate variables $y_n, n \in \mathcal{N}$. We directly exhibit a serie of matrices whose psd insure that $z \leq \prod_n w_n$. The result for the NBS relies on the following:

Proposition 15. *Let z and $w_1, \ldots, w_n$ be real positive numbers. Then, using SDP constraints, it is possible to bound these numbers by the relation:*

$$z^{2^{\lceil \log_2(N) \rceil}} \leq \prod_{i=1}^{N} w_i.$$

Proof: Let p be the smallest integer such that $2^p \geq N$. We construct a family of real positive variables $z_{i2^k+1,(i+1)2^k}$ with $1 \leq k \leq p$ and $i \in \{0, \ldots, 2^{p-k}-1\}$ satisfying the constraints expressed by the following $2^p - 1$ matrices:

$$\begin{pmatrix} z_{2i2^{k-1}+1,(2i+1)2^{k-1}} & z_{i2^k+1,(i+1)2^k} \\ z_{i2^k+1,(i+1)2^k} & z_{(2i+1)2^{k-1}+1,(2i+2)2^{k-1}} \end{pmatrix} \succcurlyeq 0,$$

where we denote $z_{jj} = w_j$ for $j \in \{1, \ldots, n\}$ and $z = z_{1,2^p}$. Then we obtain constraints of the forms $z_{1,2}^2 \leq w_1 w_2$, $z_{3,4}^2 \leq y_3 y_4$, $z_{1,4}^2 \leq z_{1,2} z_{3,4}$ and finally $z_{1,2^p}^{2^p} \leq \prod_i w_i$. ∎

To solve the NBS problem, we propose to add artificial connections so that the total number of connections is of the form 2^p and use the previous proposition. We also create fictitious links that these connections will use so that they do not modify the share of the others. It is important to bound the bandwidth allocated to these connections so that the share obtained by SDP does not grow without control that would provoke an error. As these fictitious connections use their own link, no matter the value of their upper bound PR_i, we will obtain $x_i = PR_i$.

The following proposition shows that the blocs of the matrix are of size at most 2 and that their number remains reasonable (of the order of $\mathcal{O}(\mathcal{N} + \mathcal{L})$). The SDP approach can hence offer simple and fast solutions for dimensioning purposes or for studying existing system, and is suitable for the study of large networks.

Proposition 16. *We stress that the proposed construction will lead to at most $6|\mathcal{N}| - 7$ variables, $4|\mathcal{N}| - 5$ blocks of size 2, and $4|\mathcal{N}| + |\mathcal{L}| - 4$ blocks of size 1.*

Proof: Let p be the smallest integer such that $2^p \geq |\mathcal{L}|$ and $q = 2^p$. Then, one can check that, our method will lead to: $q + \sum_{i=0}^{p} q2^{-i} = 3.2^p - 1$ variables, $q + \sum_{i=1}^{p} q2^{-i} = 2^{p+1} - 1$ blocks of size 2 and $l + 2.2^p = 2^{p+1} + l$ blocks of size 1. In the worst case, we have $|\mathcal{L}| = 2^{p-1} + 1$, hence the result. ∎

15.8.4 A simple example of NBS computation with fixed routing

We finally illustrate in this sub-section the previous results with the construction of the matrix for the computation of the NBS on a simple network. We consider a network with $|\mathcal{L}| = 4$ links and $|\mathcal{N}| = 3$ connections. The routing matrix is:

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Recall that element $a_{i,j}$ equals 1 if and only if connection j uses link i. In our SDP program, we add an artificial connection so that the total number of connections has the form 2^p with $p \in \mathbb{N}$ as explained in 15.8.3. We suppose that this extra connection uses its own link and therefore does not modify the NBS of our problem. A possible network associated with this problem is represented in Figure 15.8.

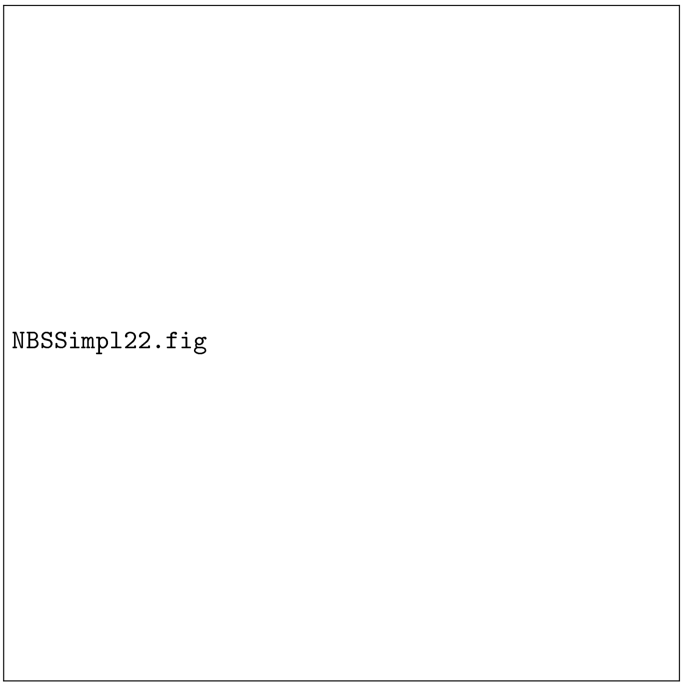


Figure 15.8: Simple computation of the NBS.

The first four blocs of the matrix link the variables x_n with their utility w_n:

$$\mathrm{MAT}_{1,n} = \begin{pmatrix} -\frac{w_n - c_n}{a_n} & x_n - b_n \\ x_n - b_n & 1 \end{pmatrix}$$

The following matrices link the w_i variables together to obtain a single variable that SDP will maximize (from Proposition 15).

$$\mathrm{MAT}_{2,1} = \begin{pmatrix} w_1 & y_{1,2} \\ y_{1,2} & w_2 \end{pmatrix}, \mathrm{MAT}_{2,2} = \begin{pmatrix} w_3 & y_{3,4} \\ y_{3,4} & w_4 \end{pmatrix}, \mathrm{MAT}_{2,3} = \begin{pmatrix} y_{1,2} & y_{1,4} \\ y_{1,4} & y_{3,4} \end{pmatrix}.$$

The positiveness of these matrices implies that $(y_{1,4})^4 \le (y_{1,2})^2 \cdot (y_{3,4})^2 \le w_1 \cdot w_2 \cdot w_3 \cdot w_4$. Then, maximizing the single variable $y_{1,4}$ will lead to the required optimization of $\prod_n (c_n - a_n(x_n - b_n)^2)$.

We incorporate the (linear) constraints of the problem. The constraints $(Ax)_l \le C_l$ lead to the declaration of L matrices that are, in our example:

$$\mathrm{MAT}_{3,1} = (C_1 - (x_1 + x_3)), \quad \mathrm{MAT}_{3,2} = (C_2 - x_2), \quad \mathrm{MAT}_{3,3} = (C_3 - (x_2 + x_3))$$
$$\text{and } \mathrm{MAT}_{3,4} = (C_4 - x_1).$$

Finally, the constraints $x_i \le PR_i$ and $x_i \ge MR_i$ are reflected by 8 scalar matrices:

$$\mathrm{MAT}_{3,4+n} = (PR_n - x_n), \quad \text{and} \quad \mathrm{MAT}_{3,8+n} = (x_n - MR_n), \quad 1 \le n \le 4.$$

We can notice that the values PR_4 and MR_4 corresponding to the artificial connection are not important since the connection is independent of the others and SDP will give $x_4 = PR_4$.

The entries we should give to the SDP algorithm are the matrix obtained by concatenation of the blocs we described and the vector L. As we want to maximize $y_{1,4}$, the vector should be of the form: $L = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, -1)$ (or an equivalent one if we renumber the variable in a different order). We can notice that we have in this simple example 11 variables: 4 allocation variables x_n, 4 utility variables w_n and 3 intermediate variables $y_{1,2}$, $y_{3,4}$ and $y_{1,4}$.

We present in the last section some numerical results obtained with the csdp solver. The matrices we used were obtained according to the method we previously explained.

15.9 Numerical experiments

We implemented the SDP approach using a Matlab program on a SUN ULTRA 1 computer to obtain the fair shares. We first tested our program on the same linear network example for which we had explicit expressions, and the results completely agreed. We then considered two more complex networks which we describe below. The computation time (including the display part) in both cases was less than a minute.

The utility functions are chosen identical for each connection (although the program can handle different parameters without increasing the complexity) and are $MR = 10$, $PR = 80$, $a = 1/490$, $b = 745$ and $c = 1102.5$. The bandwidth parameters and the allocations are given in percentage of the total capacity of the link. For each network, we present a figure showing the set of links and a figure representing the bandwidth allocated to each connection. Finally, in the connection-aware case, the links have the same capacity $C = 100$, while in the case of network-aware optimization experiments, the link sizes and the connections requirements are different one from another.

15.9.1 NBS for a small network with fixed routing

We consider the network represented in Figure 15.9. It contains $|\mathcal{L}| = 10$ links and $|\mathcal{N}| = 11$ connections. The routing is characterized by matrix A:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

The SDP solution is given in Figure 15.10, in which each connection is represented by a line whose gray scale represent its allocation. We can note that the SDP formulation required to introduce 36 variables and the SDP matrix was of size 104 (31 matrices of size 2, and 42 of size 1).

We can note that all the links cannot be saturated simultaneously even although the solution is Pareto optimal. This is for instance the case of links 1, 3 and 4. Moreover, we can distinguish 2 independent systems:

- The one formed by the links 1 to 4 and used by exclusively by connections 2, 8 and 9,
- The system that consists of links 5 to 11, used by the other connections.

Note that in the first system, only one link is saturated. This link being used by the three connections, each of them received one third of its capacity (since they have the same utility function).

The second system is more complex and we cannot determine, without the help of computers the different allocations of the connections. Still, the numerical results agree with our expectations since connection 6, using only one link, obtain the highest allocation (almost 63% of the link capacity), whereas connection 10, crossing the maximum number of links, obtain the smaller allocation (17% of the link capacity). In the case of max-min optimization these differences would have been, of course, smaller.

Finally, let us recall that the critical element in the allocation of a connection is the set of saturated links and not the total number of links crossed by a connection. Thus, in our example although connection 3 crosses more links than connection 11, it receives the same allocation. Indeed, the extra link it goes through is link 7, which is not saturated. The same observation holds for connections 1 and 4.

15.9.2 The COST network

We then considered the COST network represented in Figure 15.11. It contains 11 nodes representing major European cities. We present two optimization schemes, corresponding respectively rezosimp2.fig

Figure 15.9: Small network with fixed routing.

simplenetresults.fig

Figure 15.10: First network: solution

to the connection-aware optimization with fixed routing and to the network-aware optimization with fractional routing.

NBS in the case of connection-aware optimization with fixed routing. We considered in the simulation the 30 connections having the higher forecast demands[9]. The solution obtained in the case of the NBS is given in Figure 15.12 and the results are summarized in Table 15.2. The routing was chosen arbitrarily in order to minimize the number of links crossed by each connection.

Pa - Mi	47.34	Mi - Vi	63.00	Pr - Co	80.00	Zu - Pa - Lo	21.87	Zu - Pr - Be	50.00
Pa - Lo	33.93	Pr - Be	50.00	Be - Am	27.11	Pa - Zu - Vi	25.48	Zu - Lu - Br	35.79
Pa - Be	80.00	Lo - Br	80.00	Zu - Vi	55.06	Be - Am - Lu	27.11	Lo - Am - Be	23.73
Pa - Br	43.66	Lo - Am	76.27	Vi - Be	63.00	Mi - Vi - Be	37.00	Pa - Br - Am	28.42
Pa - Zu	33.19	Am - Br	49.54	Mi - Pa - Br	27.93	Be - Am - Br	22.04	Mi - Zu - Lu - Am	28.42
Mi - Zu	71.58	Co - Be	80.00	Mi - Pa - Lo	24.74	Zu - Lu - Am	35.79	Vi - Zu - Pa - Lo	19.46

Legend: **Co** : Copenhagen, **Be**: Berlin, **Am**: Amsterdam, **Lo**: London, **Br**: Brussels, **Pa**: Paris, **Lu**: Luxembourg, **Pr**: Prague, **Zu**: Zurich, **Vi**: Vienna, **Mi**: Milan.

Table 15.2: Bandwidth allocation in COST networks.

The solution involved adding extra 65 intermediate variables, and the size of the psd matrix was 215 (63 matrices of size 2, and 89 matrices of size 1). As in the previous example, the connections having the higher allocations are represented with darker colors. We can again note that the connections using the smaller number of (saturated) links received more bandwidth. We can even point out that four connections (Copenhagen-Prague, Copenhagen-Berlin, London-Brussels and Paris-Berlin) are alone in their respective link, and are therefore allocated their

[9] We withdrew the connections whose demands would be lower than 2.5 Gb/s according to experiments dating from 1993.

Figure 15.11: COST network: links.

maximum demand, that is to say 80% of the link capacity. Those links are thus not saturated. The allocations of connections Berlin-Prague and Berlin-Prague-Zurich are equal and are 50% of the link capacity because they compete for the same bottleneck link (the reason is therefore similar that for connections 1 and 4 of the previous network).

Network-aware optimization We then considered the case of network-aware optimization and fractional routing. In this example we added 3 links: London-Copenhagen, Amsterdam-Copenhagen and Luxembourg-Prague to increase the possible routing strategies. Moreover we got interested in comparing different values of α. We present here the results obtained for $\alpha = 0.5$ and $\alpha = 5$.

The link capacity is given in arbitrary units and we considered the 110 possible demands (corresponding to all the possible pairs source-destination). The demands are different from one to another, as illustrated in Table 15.3. Here, although the problem size is much higher than the previous example (due to the increase of the number of links and connections), the computation time remained lower than one minute. The link usages (in percentage) for $\alpha = 0.5$ and $\alpha = 5$ are given in the Tables 15.4 and 15.5 respectively and graphically represented in Figures 15.13 and 15.14.

Pa - Mi	5	Pa - Zu	6	Pa - Lu	1	Am - Lu	1	Lo - Pr	1	Pr - Mi	1	Pr - Be	2	Zu - Lo	3
Pa - Lo	10	Pa - Vi	2	Pa - Am	5	Pa - Co	1	Lo - Br	4	Lo - Am	5	Pr - Zu	1	Zu - Br	6
Pa - Pr	1	Pa - Be	11	Mi - Am	2	Mi - Lu	1	Lo - Be	8	Lo - Co	1	Pr - Br	1	Pr - Lu	1
Pa - Br	6	Mi - Be	9	Mi - Co	1	Zu - Lu	1	Am - Br	4	Br - Co	1	Lu - Lo	1	Pr - Am	1
Mi - Vi	3	Mi - Br	2	Am - Co	1	Lu - Br	1	Be - Br	6	Co - Be	3	Vi - Co	1	Vi - Lo	2
Mi - Zu	6	Mi - Lo	3	Zu - Am	3	Zu - Co	1	Vi - Br	1	Be - Lu	2	Pr - Vi	1		
Zu - Vi	3	Zu - Be	11	Pr - Co	1	Be - Am	8	Vi - Be	9	Vi - Lu	1	Vi - Am	1		

Table 15.3: COST network: demands

Link	BP	Link	BP	Link	BP	Link	BP	Link	BP	Link	BP	Link	BP
Pa - Mi	57	Zu - Pr	87	Vi - Be	96	Am - Lo	61	Mi - Zu	61	Pa - Br	68	Mi - Br	73
Pa - Zu	68	Pr - Vi	55	Be - Am	92	Am - Co	29	Mi - Vi	81	Pr - Lu	32	Am - Lu	62
Pa - Be	97	Pr - Be	93	Be - Co	81	Am - Br	90	Zu - Lu	73	Lu - Br	61	Br - Lo	62
Pa - Lo	91	Zu - Vi	41	Pr - Co	61	Lo - Co	87						

Table 15.4: Results for $\alpha = 0.5$

We observe that, when the parameter α grows, we tend to a more homogenous coloration of the links, which agrees with theoretical results. When α grows, the allocations are more "equaled" and

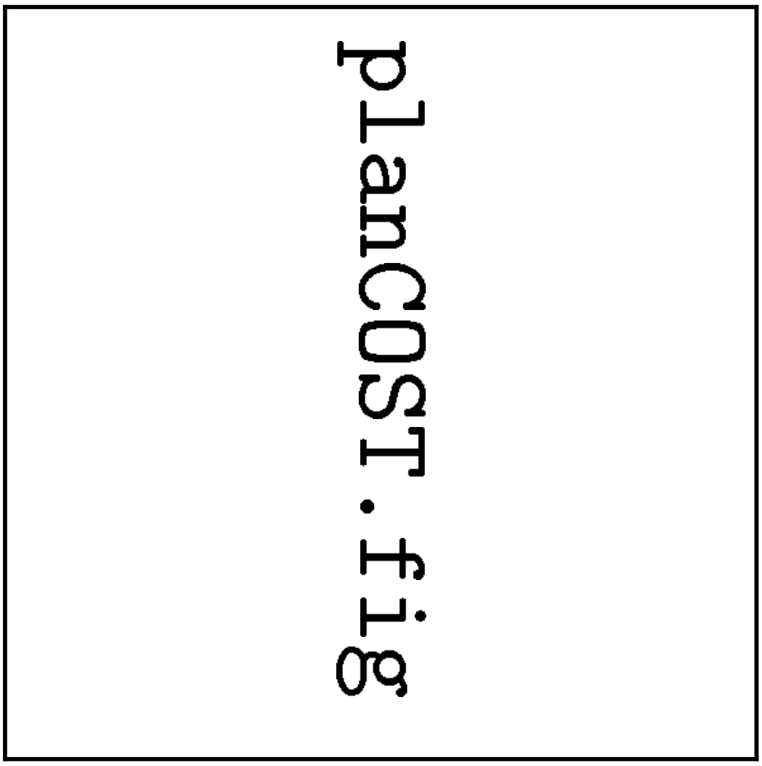


Figure 15.12: COST network: NBS allocation.

Link	BP	Link	BP	Link	BP	Link	BP	Link	BP	Link	BP	Link	BP
Pa - Mi	52	Mi - Zu	73	Zu - Lu	76	Am - Lo	73	Mi - Br	76	Pa - Lo	78	Pr - Co	74
Pa - Zu	73	Pr - Vi	76	Be - Am	92	Am - Co	47	Vi - Be	92	Am - Br	77	Lo - Co	73
Pa - Be	92	Pr - Be	92	Be - Co	92	Lu - Br	73	Br - Lo	74	Mi - Vi	73	Pa - Br	76
Zu - Pr	76	Pr - Lu	47	Am - Lu	74	Zu - Vi	47						

Table 15.5: Results for $\alpha = 5$

we thus observe a more homogeneous use of the links. Accordingly, the use of the more saturated link decrease from 97.36% to 92%.

When α grows from 0.5 to 5, the average link usage increases (from 70.38% to 73.83%). This can seem surprising, since the allocated bandwidth remain constant, as well as the link capacity. In fact, when α grows, the traffic becomes more split between the different possible routes in order make the link usage more homogeneous. Therefore the average number of links used by a connection increase as well and so does the network usage.

It is also interesting to note that the usage of some links, like the Berlin-Copenhagen, have increased almost to their maximum usage for the solution $\alpha = 5$, supposedly more "fair" ! This shows that the search for fairness in a global scale sometimes lead some users to critical situations while their remain in their bound in cases that are less fair.

15.10 Conclusion

We have applied in this paper the NBS approach for bandwidth allocation that is sensitive to the utilities of connections. We have studied some of the characteristics of these concepts, and showed that they are indeed more suitable for applications that have concave utility. We proposed a simple parametrization of the concavity of the utility function using quadratic functions. We finally proposed some computational approaches that allows us to handle large networks: a Lagrangian approach and a novel approach based on SDP.

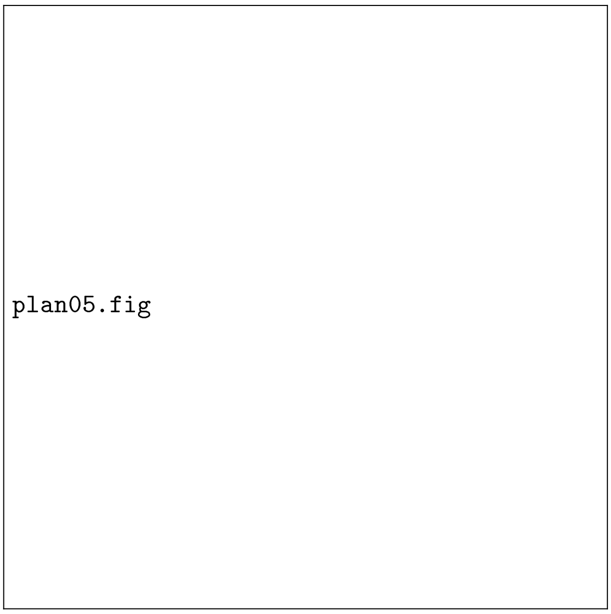


Figure 15.13: Bandwidth allocation for COST network, $\alpha = 0.5$.

The SDP approach has the advantage that it is simple to implement using any general SDP software. Furthermore additional conditions (for instance those linked to integer programming, or other telecommunication requests) can be introduced without requiring any study on the stability or the convergence of the algorithm. This is a clear improvement compared to many other specific methods, and in particular to the iterative ones that have been proposed for solving fairness problems.

Also, we considered a fairness approach that takes into account not just the assigned throughput of each connection but instead the utility that the assigned throughput represents. This type of fairness concept agrees with the game theoretic definition of fairness given by NBS and is an interesting generalization of the fairness criteria that have been used so far in the telecommunication context.

.1 Proof of Proposition 3

We first prove the following lemma:

Lemma 5. *As $|\mathcal{L}|$ grows to infinity, the only possible limits x_{lim} of the bandwidth assigned to connection 0 are MR_0 and $C_0 - b_1$.*

Proof: Since x is bounded (from $MR_0 \leq x \leq PR_0$), the left part of (15.8) is bounded too. Therefore, the limit x_{lim}, if any, is such that $a_1(C_0 - x - b_1) \cdot f_0(x) = 0$.
We want to determine which of the three solutions are possible, that is to say belonging to the set $[MR_0, PR_0]$. From (15.7) we have $PR_0 + PR_1 > C$. As $b_1 > PR_1$ (from (15.5)), then $C - b_1 < PR_0$. Therefore the solution $x_{lim} = C - b_1$ is acceptable if $C - b_1 \geq MR_0$. The two other solutions are the zeros of f_0. By definition the larger one is strictly greater than PR_0 and therefore not acceptable. ∎

The second lemma compares the possible limits.

Lemma 6. *Let f and g be defined as in (15.8). Then $g(C-b_1) \geq g(MR_0)$ and $h(C-b_1) \geq h(MR_0)$.*

Proof: We have $g(MR_0) = \dfrac{|\mathcal{L}|}{1-\alpha} f_1(C - MR_0)^{1-\alpha}$ and plan5.fig

Figure 15.14: Bandwidth allocation for COST network, $\alpha = 5$.

$g(C - b_1) = \frac{1}{1-\alpha}\left[f_0(C - b_1)^{1-\alpha} + |\mathcal{L}|(f_1(b_1))^{1-\alpha}\right].$
From the definition of b_1 we have $f_1(b_1) \geq f_1(C - MR_0)$. Moreover, if $C - b_1 \geq MR_0$ then $f_0(C - b_1) \geq 0$ and therefore $g(C - b_1) \geq g(MR_0)$.
If $\alpha = 1$, then $h(MR_0) = 0$ and $h(C - b_1) = f_0(C - b_1)(f_1(b_1))^{\mathcal{N}} > 0$. ∎

We can finally prove Proposition 3. We have seen that for any value of α, MR_0 and $C - b_1$ are the only two possible limits. Moreover, if $C - b_1 \geq MR_0$ then $g(C - b_1) \geq g(MR_0)$ and $h(C - b_1) \geq h(MR_0)$. Therefore, the limit of the allocation, if any, is $\max(C - b_1, MR_0)$.
We can show that this value is indeed the limit of the allocation. We detail here the case of $\alpha = 1$. The case of $\alpha \neq 1$ being similar. If $C - b_1 > MR_0$, then $f_1(C - Id)$ is maximal for $C - b_1$ (by definition of b_1) and $f_0(C - b_1) > 0$ (since $C - b_1 > MR_0$), then, $lim_{|\mathcal{L}| \to \infty} f_0(x)(f_1(C - x))^{|\mathcal{L}|} = C - b_1$.
Suppose now that $C - b_1 \leq MR_0$. $f_1(C - \mathrm{Id})$ is a parabola those maximum is for $C - b_1 \leq MR_0$ and that is null in MR_0. Therefore the function is decreasing in $[b_1, +\infty]$, and in particular in $[C - MR_0, C - PR_0]$ (since $C - b_1 > MR_0$). Therefore its maximum is achieved in $C - MR_0$. As f_0 is positive on the set $[MR_0, PR_0]$ then $x_{lim} = MR_0$.

Appendix A

Coalition games and the multicast problem

A.1 Introduction

This Chapter is based on a joined paper by Chandramani [?] and Altman.

A.1.1 The Multicast Problem

Consider a network with one source or one base station (BS). N users participate in a multicast session in which each one receives the same content from the source. Users are non-homogeneous: Transmission of the content is associated with some costly resource (power, bandwidth etc). There is some amount of resource that a user requests associated with the quality of the content it wishes to receive as well as to the network conditions of that user. It is assumed that if a given amount of resource is used for transmission, then all users with a request not exceeding this amount are satisfied. In order to satisfy all users, the source needs to allocate the amount that corresponds to the largest request. Some examples are:

- Power control: consider a cell with N mobiles. A mobile uses a service that needs some given power level at the reception. As the locations of mobiles differ, the channel gain differs from one mobile to the other. Thus the BS has to transmit at a power level that ensures that the mobile with the worst channel conditions receives the signal with the requested power. For example, if the channel gain between the base station located at X_{bs} and mobile i, located at X_i, is determined by the distance and the path loss constant α, i.e. $p_r = Pd(X_{bs}, X_i)^{-\alpha}$ then the BS has to transmit at a power of $P = \theta \max_i d(X_{bs}, X_i)^{\alpha}$ in order to ensure that all mobiles receive at a power of at least θ.

- Power control with quality of service constraints: We consider again a BS and N mobiles. The bit error rate (BER) at the reception is known to be a function of the signal to noise ratio and of the modulation type of the signal. Each user may wish to receive at a different quality in terms of BER. Therefore the power needed to satisfy all users is not necessarily the one corresponding to the mobile who is the furthest away.

- Hierarchical multicast. Assume that a hierarchical coding is used. The content is coded and transmitted over several carriers. A user can decide what level of details he wishes to have; this translates to the decision of how many carriers the user wishes to receive. The more carriers one receives, the better is the quality of the received signal after decoding. If mobile i subscribes to j_i carriers, then in order to satisfy all users, the source has to use $\max_{i=1}^{N} j_i$ carriers.

A.1.2 The structure of the chapter

The first question we address in this chapter is how to split the cost for establishing the multicast session among the users. In the situations described above it seems clear that the global benefit increases in the size of the multicast group. Nonetheless, depending on the splitting rule, users may join the multicast session or may prefer not to join it. Indeed, it is assumed that a user or a set of users that are not satisfied with the way the cost is split, can form an alternative multicast group that would perhaps require together less resources and may thus be more beneficial for these. We look for rules for sharing the multicast cost that will make it advantageous for all participants to form one large multicast group, which is called then the "grand coalition". We use tools from cooperative game theory where multicast groups are viewed as coalitions. The set of such rules is called **the core**.

We next consider explicitly an alternative unicast source that each mobile can connect to at some constant cost. We loose the property that we had before where we always had splitting rules that made the grand coalition appealing to everyone. One may view this problem as a hierarchical non-cooperative association problem: given the leader's rules (the rules for splitting the cost among the participants of the multicast group), each player has the choice to join the multicast group or a dedicated unicast one. We study the impact of the cost sharing rule in the multicast group on the number of participants in it (which we call capacity) and on its geographical size (which we call coverage). (Note: One can formulate this problem as one with multiple coalition structure, as defined in [73, p. 44, section 3.8].) We study the impact of information on the performance. We discover a paradoxical behavior in which the performances improves by providing less information to mobiles.

A.1.3 Related Work

The cost structure in our problem is identical to that proposed by Littlechild and Owen [74, 75] in the context of *Aircraft landing fees*. Thomson [76] provides a survey on cost allocation for the airport problem.

Myerson [77] developed the theory of large Poisson games, and in particular, proved the existence of equilibria in such games. Under the setup there, utility of a player depends on the aggregate action profile of the whole population, and not on the type-wise action profile. In our setup the cost of a user depends on the type-wise action profile of the population, where type of a user can be identified as its location.

Penna and Ventre [78] and Bilo et al. [79] study the problem of sharing the cost of multicast transmission in a wireless network. In another paper, Bilo et al. [80] frame the selfish nature of users as a noncooperative game among them, for several given cost allocation methods.

A.2 Coalition Game Preliminaries

We begin by defining the coalition game. A cooperative cost game [81] is a pair $(\mathcal{N}, c)$ where $\mathcal{N} := \{1, \ldots, N\}$ denotes the set of players and $c : 2^{\mathcal{N}} \to \mathbb{R}$ is the cost function For any nonempty coalition $\mathcal{S}$, $c(\mathcal{S}S)$ is the minimal cost incurred if players in $\mathcal{S}$ work together to serve their purposes; $c(\emptyset) = 0$. A cooperative game is called *concave* if the cost function is *sub-modular* (the precise definition is delayed to Eqs (A.3)).

A cost allocation $\mathbf{q} \in \mathbb{R}^{\mathcal{N}}$ charges cost q_i to player i. An allocation $\mathbf{q}$ is called *efficient* if $\sum_{i\in\mathcal{N}} q_i = c(\mathcal{N})$. An efficient allocation $\mathbf{q}$ is called an *imputation* if $q_i \le c(\{i\})$ for all $i \in \mathcal{N}$.

The core: The core, $\mathcal{C}$, of the game is defined as follows

$$\begin{aligned} \mathcal{C} = \{\mathbf{q} \in \mathbb{R}^{\mathcal{N}} \quad : \quad & \sum_{i\in\mathcal{N}} q_i = c(\mathcal{N}), \\ & \sum_{i\in\mathcal{S}} q_i \le c(\mathcal{S}), \ \forall\, \mathcal{S} \subset \mathcal{N}\} \end{aligned} \tag{A.1}$$

The core of a concave cooperative game is nonempty [82].

Next we state a number of appealing rules for cost allocation.

- **Shapley value:** For any i, and $\mathcal{S} \subset \mathcal{N}$ such that $i \notin \mathcal{S}$, let $\Delta_i(\mathcal{S}) = c(\mathcal{S} \cup \{i\}) - c(\mathcal{S})$. The Shapley value is the cost allocation $\mathbf{q}$ for which

$$q_i = \frac{1}{n!} \sum_{U \in \mathcal{U}} \Delta_i(\mathcal{S}_i(U)), \tag{A.2}$$

  where $\mathcal{U}$ is the set of all orderings of $\mathcal{N}$, and $\mathcal{S}_i(U)$ is the set of players preceding i in ordering U. The Shapley value of a concave cooperative game lies in the core.

- **Nucleolus:** The *excess* of a coalition $\mathcal{S}$ under an imputation $\mathbf{q}$ is $e_S(\mathbf{q}) = \sum_{i \in \mathcal{S}} q_i - v(\mathcal{S})$; this is a measure of dissatisfaction of $\mathbf{S}$ under $\mathbf{q}$. Let $E(\mathbf{q}) = (e_S(\mathbf{q}), S \in 2^{\mathcal{N}})$ be the vector of excesses arranged in monotonically increasing order. The *nucleolus* is the set of imputations $\mathbf{q}$ for which the vector $E(\mathbf{q})$ is lexicographically minimal.

  Nucleolus is a singleton and belongs to the *core* whenever the latter is nonempty.

- **Egalitarian Allocation:** The egalitarian allocation for cooperative games was introduced by Dutta and Ray [83]. It is unique whenever it exists, and it always exists and lies in the core for concave cost games. The following characterization applies to such games only.

  The egalitarian allocation is the element of core which Lorentz dominates all other core allocations.

  Remark 15. *A min-max fair allocation for the cooperative cost game is defined as follows. For $\mathbf{q} \in \mathbb{R}^{\mathcal{N}}$, define $\bar{\mathbf{q}}$ to be the vector obtained by arranging the components of $\mathbf{q}$ in decreasing order. Further define*

$$\bar{\mathcal{C}} = \{\bar{\mathbf{q}} : \mathbf{q} \in \mathcal{C}\}.$$

  Then $\mathbf{q} \in \mathcal{C}$ is a min-max fair allocation if and only if $\bar{\mathbf{q}}$ is lexicographically minimal in $\bar{\mathcal{C}}$.

  A max-min fair allocation is also defined similarly. Jain and Vazirani[84] show that for concave games, there exists a unique cost allocation which is min-max fair as well as max-min fair, and which coincides with the egalitarian allocation.

A.3 System Model

A.3.1 Network and Communication Model

We consider a wireless network with a base station (BS) and a set $\mathcal{N} = \{1, \ldots, N\}$ of users. We shall assume N to be either a known constant (in the case of perfect information), or a Poisson random variable of rate λ (in the framework of imperfect information).

Both the radio channel varies from one mobile to another, as well as the required QoS level. Mobile i requires transmission at power p_i in order to meet its QoS needs. We assume p_is to be independently identically distributed (i. i. d.) random variables with distribution $G(p)$ (density $g(p) := G'(p)$).

We assume that any subset S of users can subscribe for a multicast session. The BS then broadcasts information (say, a radio channel) with the minimum power p that guarantees that all mobiles in S receive a satisfactory level of quality of service. Evidently $p = \max\{p_i : i \in S\}$.

Assume that there is a cost $f(p)$ per time unit during which the BS transmits at power p; this cost has to be shared by all the mobiles in the multicast group. f is assumed to be increasing in its argument.

We also assume that every user has yet another option, that of using a dedicated connection using some other technology, at a cost V.

A.3.2 Cost Sharing Models

Let us assume that the set of mobiles in multicast group is N. We index the mobiles such that such that p_i is increasing in i.

A cost sharing mechanism is a map $\mathbf{q} : \mathbb{R}^{\mathcal{N}} \to \mathbb{R}^{\mathcal{N}}$ with elements $q_j, j \in \mathcal{N}$; q_j is the cost share of user j. We study the cost sharing mechanisms that satisfy the following economical constraints.

Budget-balance A cost sharing mechanism is called *budget-balanced* if users pay exactly the total cost of the service.

$$\sum_{j=1}^{N} q_j = f(p_N).$$

cross-monotonicity A cost sharing mechanism is called *cross-monotonic* if each user's cost decreases as the service set expands.

efficiency An *efficient* cost sharing mechanism is one that maximizes net social utility.

Remark 16. *A cost sharing mechanism is called strategy-proof if revealing true utilities is a dominant strategy for each user. However, in the cost sharing problem studied here all the users are assumed to have equal utility which is known.*

The cost sharing problem can be formulated as a cooperative cost game $(\mathcal{N}, c)$. Here $c : 2^{\mathcal{N}} \to \mathbb{R}$, for a coalition $\mathcal{S} \subset \mathcal{N}$, gives the cost to support communication to all the users in $\mathcal{S}$, i.e., $c(\mathcal{S}) = \max\{f(p_i) : i \in \mathcal{S}\}$.

Now consider two coalitions $S_1, S_2 \subseteq \mathcal{N}$. Observe that

$$\begin{aligned} c(S_1 \cup S_2) &= \max\{c(S_1), c(S_2)\} && \text{(A.3)} \\ \text{and } c(S_1 \cap S_2) &\leq \min\{c(S_1), c(S_2)\}. && \text{(A.4)} \end{aligned}$$

Hence,

$$c(S_1 \cup S_2) + c(S_1 \cap S_2) \leq c(S_1) + c(S_2),$$

i.e., the cost function is submodular. This implies the following.

Theorem 21. *(i) The core of the cost allocation game is nonempty.*
(ii) The Shapley value lies in the core.
(iii) The egalitarian allocation lies in the core and is min-max (also max-min) fair.

The core of the game can be expressed as

$$\Big\{\mathbf{q} \in \mathbb{R}^{\mathcal{N}} : \sum_{i \in \mathcal{S}} q_i \leq f(p_{\max \mathcal{S}}), \mathcal{S} \subset \mathcal{N}, \sum_{i \in \mathcal{N}} q_i \leq f(p_N)\Big\},$$

where $\max(S) := \max i : i \in (S)$. We make the following observations

1. All the cost allocation in the core are nonnegative; if $q_i < 0$, $\mathbf{q}$ can not satisfy the constraint corresponding to subset $\mathcal{N} \backslash \{i\}$.

2. The constraint $\sum_{i=1}^{j} q_i \leq f(p_j)$ makes the constraints corresponding to the subsets $S \subsetneq \{1, \ldots, j\}$ redundant.

In view of these, the core can be rewritten as

$$\Big\{q \in \mathbb{R}_+^N : \sum_{i=1}^{j} q_i \leq f(p_j),\ 1 \leq j \leq N, \sum_{i=1}^{N} q_i = f(p_N)\Big\}$$

A budget-balanced cost sharing mechanism is cross-monotonic only if it belongs to the core of the associated cooperative cost game. Hence we focus on cost allocations from the core. Following criteria can be used.

1. **Highest cost allocation (HCA):** The user requiring the highest power, N, pays the whole cost. Of course $(0, \ldots, f(p_N))$ is in the core.

2. **Incremental cost allocation (ICA):** A more fair cost allocation is where user i pays $f(p_i) - f(p_{i-1})$; $f(p_0) := 0$. We call it incremental cost allocation.

3. **Shapley value (SV):** Following [74], the cost q_i for player i is given by

$$q_i = \sum_{j=1}^{i} \frac{f(p_j) - f(p_{j-1})}{N + 1 - j}$$

4. **Nucleolus (NS):** The following algorithm for calculating the nucleolus is given by Littlechild [75] in the context of the *airport cost game.*

   Define $i_0 = r_0 = 0$. For $k \geq 1$, iteratively define

$$r_k = \min_{i_{k-1}+1, \ldots, n-1} \left\{ \frac{f(p_i) - f(p_{i_{k-1}}) + r_{k-1}}{i - i_{k-1} + 1} \right\},$$

   and i_k as the largest value of i for which the minimum is attained in the above expression. Continue this until $k = k'$ where $i_{k'} = N - 1$. The nucleolus of the game, q, is given by

$$\begin{aligned} q_i &= r_k, \ i_{k-1} < i \leq i_k, \ k = 1, \ldots, k' \\ q_N &= f(p_N) - f(p_{N-1}) + r_{k'} \end{aligned}$$

5. **Egalitarian allocation (EA):** The egalitarian allocation for our cost sharing problem can be computed by applying the following algorithm [85].

   Define $i_0 = r_0 = 0$. For $k \geq 1$, iteratively define

$$r_k = \min_{i_{k-1}+1, \ldots, n} \left\{ \frac{f(p_i) - f(p_{i_{k-1}})}{i - i_{k-1}} \right\},$$

   and i_k as the largest value of i for which the minimum is attained in the above expression. Continue this until $k = k'$ where $i_{k'} = N$. The max-min fair allocation, q, is given by

$$q_i = r_k, \ i_{k-1} < i \leq i_k, \ k = 1, \ldots, k'$$

Cross-monotonicity of the above cost allocations: Evidently *HCA* is cross-monotonic. Shapley values is well known to be cross-monotonic (see Moulin[86]). Dutta [87] showed that the egalitarian allocation in concave games is also cross-monotonic. Sonmez [88] showed that the nucleolus of a generic concave cost allocation game need not be cross-monotonic; however he also proved the cross-monotonicity of nucleolus for the airport game which has identical formulation as ours.

We observe also the following monotonicity property of cost allocations *HCA, SV, NS* and *EA*. *"For any two users i, j such that $p_j > p_i$, $q_j \geq q_i$."*

The expression for nucleolus has the similar form as that of max-min fair allocation. Hence, in the following we analyze max-min fair allocation but do not discuss nucleolus.

A.4 Non-cooperative Coalition Formation game: perfect information

Each player independently decides whether to join the multicast group or not. Recall that, a player if it does not use the multicast service, bears a cost V. We formulate the decision problem as a noncooperative game. Assume that all users know about the resources requested by all other users in the network.

Definition 29. *An equilibrium is a multicast subset $\mathcal{M} \subset \mathcal{N}$ of users, such that the cost share of each one in that set is not greater than V, and the cost of each user not in the set would not be smaller than V if it joined the set.*

Let p_i be the amount of resources required by player i. Let $p_{\mathcal{M}}$ denote the set $p_i, i \in \mathcal{M}$ and let q_i be the cost share of player i. It is given as some function h_i that depends on the set $\mathcal{M}$ and on the resource requests of each player in the set $\mathcal{M}$. With this notation, the equilibrium is characterized by the following conditions: $h_i(\mathcal{M}, p_{\mathcal{M}}) \leq V$ for all $i \in \mathcal{M}$, and $h_i(\mathcal{M} \cup i, p_{\mathcal{M} \cup i}) \geq V$ for all $i \in \mathcal{N} \backslash \mathcal{M}$.

Remark 17. *We shall assume that the amount of resources requested by a mobile is a random realization of some probability distribution that has no mass at isolated points. For simplicity we shall focus on the case where the resource is power, and the amount required is a strictly monotone continuous increasing function of the distance of the mobile from the base station. The distance between the users and the base station are independent from user to user and are assumed to be drawn from a probability distribution that has a density. Therefore the probability of two mobiles having the same distance from the base station is zero. We shall thus assume below that all distances are different from each other. Similarly, we shall ignore the possibility (which has probability zero) that $f(p_i) = V$ for some i.*

We next provide the Nash equilibrium (NE) for the various cost sharing policies.

1. Under the rule that the user requiring highest power pays the whole cost $\mathcal{M} = \{i : f(p_i) < V\}$. This is in fact a *strong equilibrium*: it is robust not only to any deviation by a single user but to any deviation by any number of users. No coordination between users is needed for reaching this equilibrium since the best response of a user (as specified by this strategy) is independent of what other users do.

2. When users pay incremental costs,
$$\mathcal{M} = \begin{cases} \emptyset & \text{if } p_1 \geq V, \\ \{1, \ldots, j-1\} & \text{otherwise}, \end{cases} \tag{A.5}$$
where $j = \arg\min_i \{i : f(p_i) - f(p_{i-1}) \geq V\}$.

3. Shapley value: Let us define
$$b_j = \sum_{i=1}^{j} \frac{f(p_j) - f(p_{j-1})}{j + 1 - 1},$$
Then the unique NE is $\mathcal{M} = \{1, \ldots, j\}$ where j is the largest index such that $b_j \leq V$.

4. Max-min fair allocation: Define $i_0 = r_0 = 0$. For $k \geq 1$, iteratively define
$$r_k = \min_{i_{k-1}+1, \ldots, n} \left\{ \frac{f(p_i) - f(p_{i_{k-1}})}{i - i_{k-1}} \right\},$$
and i_k as the largest value of i for which the minimum is attained in the above expression. Continue this until $r_k > V$. The unique NE is $\mathcal{M} = \{1, \ldots, i_{k-1}\}$.

A.5 Static problem with incomplete information

Here we assume that any user does not know the number of other users in the network and their resource (say, power) requirements. However, the users know their own requirements. What is the equilibrium policy?

We restrict ourselves to symmetric strategies for all the users (see Myerson [77] for a discussion on this). The number of players is Poisson distributed with mean λ. *Environmental equivalence* property of Poisson games [77] ensures that from the perspective of any player, the number of other players in the game is also a Poisson random variable with the same mean λ. These arguments imply that a user's decision is a function of its power requirement, and a pure strategy equilibrium is characterized by a set such that users with power requirements in that set join the multicast group and others do not.

Let $Q \subset \mathbb{R}_+$ be such that users with power requirements in Q join the multicast group. Consider user i with power requirement p_i and let ω_{-i} denote a realization for the rest of the network. The cost share of user i is given as

$$q_i(p_i, Q, \omega_{-i}) = h_i(\mathcal{M}_Q(\omega) \cup i, p_{\mathcal{M}_Q \cup i}(\omega)),$$

where $\mathcal{M}_Q = \{i : p_i \leq Q\}$. The expected cost share of user i is $\mathbb{E}q_i(p_i, Q)$.

Definition 30. *An NE in this case is characterized by a set $Q \subset \mathbb{R}_+$ such that users with power requirements $p_i \in Q$ join the multicast group. More precisely $\mathbb{E}q_i(p_i, Q) \leq V$ if $p_i \in Q$, and $\mathbb{E}q_i(p_i, Q) > V$ if $p_i \notin Q$.*

Again consider user i, and a fixed realization of the network. We can view the function $h_i(\mathcal{M}, p_{\mathcal{M}\setminus i}, \cdot) : p_i \mapsto q_i$ as parametrized by power requirements of other users. We use $\hat{h}_i : \mathbb{R}_+ \to \mathbb{R}_+$ to denote it, i.e.,

$$\hat{h}_i(p_i) := h_i(\mathcal{M}, p_{\mathcal{M}}).$$

Following is an important observation.

Lemma 7. *Under cost allocation policies HCA, SV and EA, $\hat{h}_i$ is a monotone increasing function.*

Proof. Clearly the claim is true for *HCA*.
SV: Let us consider user i with required power p_i. Assume that its power requirement is increased to p_i'. If $p_i' \leq p_{i+1}$ then the cost share increases by

$$\hat{h}(p_i') - \hat{h}(p_i) = \frac{f(p_i') - f(p_i)}{N + 1 - i}.$$

Let us consider the case when $p_{i+1} < p_i' < p_{i+2}$. Other cases can be analyzed with a repeated application of this procedure. Now the new cost share of player i is

$$\begin{aligned}
\hat{h}(p_i') &= \sum_{j=1}^{i-1} \frac{f(p_j) - f(p_{j-1})}{N + 1 - j} \\
&\quad + \frac{f(p_{i+1}) - f(p_{i-1})}{N + 1 - i} + \frac{f(p_i') - f(p_{i+1})}{N - i} \\
&\geq \sum_{j=1}^{i} \frac{f(p_j) - f(p_{j-1})}{N + 1 - j} \\
&= \hat{h}(p_i)
\end{aligned}$$

EA: Let us revisit the algorithm used to obtain *EA* and assume $i = i_k$ for some k. Now user i increases its power requirement to p_i'. As before first consider the case when $p_i' \leq p_{i+1}$. Clearly $\hat{h}(p_i') \geq \hat{h}(p_i)$ in this case. The same holds true if $p_{i+1} < p_i' < p_{i+2}$. Similar arguments can be made in the case when $i \neq i_k$ for any k. □

Recall that we consider symmetric strategies for all the users. The following theorems show that only candidates for NEs are closed intervals containing 0.

Corollary 3. *For the cost sharing mechanisms HCA, SV and EA, under any symmetric policy Q, there exists a threshold p^* such that $\mathbb{E}q_i(p_i, Q) \leq V$ if and only if $p_i \leq p^*$.*

Corollary 4. *Under cost sharing mechanisms HCA, SV and EA, the only candidates for NEs are sets of the form $[0, p^*]$ and $[0, \infty)$.*

Following is another useful property of the above cost sharing mechanisms.

Lemma 8. *For the cost sharing mechanisms HCA, SV and EA, $\mathbb{E}q_i(p_i, [0, p])$ is monotonically increasing with p_i for $p_i \geq p$.*

Theorem 22. *For the cost sharing mechanism ICA, the only candidates for NEs are sets of the form $[0, p^*]$ and $[0, \infty)$.*

Proof. We prove the claim via contradiction. Assume that $Q = \cup_{k=1}^{K}[a_k, b_k]$ is an NE where $b_{k-1} < a_k$ for all k ($b_0 := 0$). For $0 \leq p_i \leq a_1$, user i's expected cost $f(p_i)$ will be increasing in p_i. For $b_1 \leq p_i \leq a_2$, user i's expected cost $\mathbb{E}q_i(p_i, Q)$ is

$$f(p_i) - \int_{a_1}^{b_1} \lambda f(p) g(p) \exp\Big(- \lambda \int_{p}^{b_1} g(s)\ ds\Big)\ dp.$$

In writing the above expression we have used the *decomposition property* of Poisson distribution: The number of users with power requirements in the range $[p, b_1]$ is a Poisson random variable with rate $\lambda \int_p^{b_1} g(s)\ ds$. Finally it is seen that the above expression is increasing in p_i. Similarly it can be shown that $\mathbb{E}\hat{h}_i(p_i), b_{k-1} \leq p_i \leq a_k$ is increasing for all $1 \leq k \leq K$.

Hence if all other users are using the strategy Q, Q can not be player i's best response. Thus only candidates for symmetric NEs are the threshold strategies $[0, p^*]$.

On the other hand if $\mathbb{E}q_i(p_i, [0, \infty)) \leq V$ for all p_i then $[0, \infty)$ is also an NE. □

Expressions for the NEs

We have shown that for each of the cost allocation strategies the NEs are characterized by thresholds p^*. In this section we derive expressions for the thresholds.

Theorem 23. *For the cost sharing mechanisms HCA, SV and EA, a symmetric multi-strategy $[0, p^*]$ is an NE if and only if $\mathbb{E}q_i(p^*, [0, p^*]) = V$. $[0, \infty)$ is also an NE provided $\mathbb{E}q_i(p_i, [0, \infty)) \leq V$ for all p_i.*

Proof. Suppose $\mathbb{E}q_i(p^*, [0, p^*]) = V$. Then $\mathbb{E}q_i(p_i, [0, p^*]) \leq V$ for all $p_i \leq p^*$. Also from Lemma 8, $\mathbb{E}q_i(p_i, [0, p^*]) > V$ for all $p_i > p^*$. Thus $[0, p^*]$ is indeed an NE.

Now assume $\mathbb{E}q_i(p^*, [0, p^*]) > V$. Consider user i with $p_i = p^*$. Then $[0, p^*]$ can not be an equilibrium strategy of user i. Finally assume $\mathbb{E}q_i(p^*, [0, p^*]) < V$ and denote $\epsilon := V - \mathbb{E}q_i(p^*, [0, p^*])$. Since $\mathbb{E}q_i(p_i, [0, p^*])$ is continuous and increasing in p_i for $p \geq p^*$, there exists a $\delta > 0$ such that $\mathbb{E}q_i(p_i, [0, p^*]) \leq V$ for $p_i = p^* + \delta$. Thus $[0, p^*]$ can not be an equilibrium strategy of user i.

If $\mathbb{E}q_i(p_i, [0, \infty)) \leq V$, user i's best response is to join the multicast group given that all others have joined. Hence $[0, \infty)$ is also an NE. □

HCA:

Corollary 5. *$[0, f^{-1}(V)]$ is an NE. $[0, \infty)$ is also an NE provided $f(p_i) \leq V$ for all p_i.*

Proof. For *HCA* cost sharing $\mathbb{E}q_i(p_i, [0, p_i]) = f(p_i)$. Since $f(\cdot)$ in strictly increasing, the unique solution to $\mathbb{E}q_i(p^*, [0, p^*]) = V$ is $p_i = f^{-1}(V)$. Theorem 23 proves the claim. □

ICA: Consider user i with power requirement p_i. Let us assume that all other users join the multicast group. Again using the *decomposition property* of Poisson distribution, the expected cost of user i, $\mathbb{E}q_i(p_i, [0, \infty))$ is

$$f(p_i) - \int_{0}^{p_i} \lambda f(p) g(p) \exp\Big(- \lambda \int_{p}^{p_i} g(s) ds\Big) dp.$$

Lemma 9. *$[0, p^*]$ is an NE if and only if $\mathbb{E}q_i(p_i, [0, \infty)) \leq V$ for all $p_i \leq p^*$ and $\mathbb{E}q_i(p^*, [0, \infty)) = V$. If $\mathbb{E}q_i(p_i, [0, \infty)) \leq V$ for all p_i, $[0, \infty)$ is also an NE.*

Proof. if part: Recall that, under *ICA* mechanism, user i's cost share depends on only those users that have power requirements less than p_i. Hence $\mathbb{E}\hat{h}_i(p_i, [0, p])$ are same for all $p \geq p_i$. Now, from the conditions on p^*, $\mathbb{E}\hat{h}_i(p_i, [0, p^*]) \leq V$ for all $0 \leq p_i \leq p^*$. Also, following the proof of Proposition 22, $\mathbb{E}\hat{h}_i(p_i, [0, p^*]) > V$ for all $p_i > p^*$. Hence $[0, p^*]$ is indeed an NE.

only if part: Consider a symmetric strategy $[0, p']$, and assume that there exists a $0 < p \leq p'$ such that $\mathbb{E}\hat{h}_i(p, [0, p']) > V$. Clearly $[0, p']$ can not be an equilibrium strategy of user i. Finally consider the case when $[0, p']$ is a symmetric strategy while $\mathbb{E}\hat{h}_i(p', [0, \infty)) < V$. Denote $\epsilon := V - \mathbb{E}q_i(p', [0, \infty))$. Since $f(p)$ is continuous and increasing, there exists a $\delta > 0$ such that $f(p' + \delta) - f(p') \leq \epsilon$ implying $\mathbb{E}q_i(p_i, [0, \infty)) \leq V$ for $p_i = p' + \delta$. Thus $[0, p']$ can not be an equilibrium strategy of user i. □

A.6 Static problem with no information

Next we assume that the requirement of a user is not known even to that own user. Why should a user not know its own request? The amount of resource requested may depend on the access channel quality which might not be known.

As an example, assume that receiver nodes are mobile and should decide whether to join a multicast session an hour in advance. The session is multicast by some BS. A mobile cannot predict what its distance to the BS will be an hour ahead. It only has the probability distributions of the number N of users (Poisson with mean λ), and users' distances to the BS that are i.i.d. and yield distributions $G(p)$ on the power requirements.

Consider a tagged user, say i. Given that there are n other users in contention, the expected cost share user of user i (for a given cost sharing strategy) is

$$\bar{q}_i(n) = \mathbb{E}\Big[h_i(\mathcal{M}(\omega), p_{\mathcal{M}}(\omega))\Big||\mathcal{M}(\omega)| = n + 1\Big].$$

Using the cross-monotonicity of the cost sharing strategy and a coupling argument it can be shown that $\bar{q}_i(n)$ is decreasing in n.

Now, consider the symmetric multi-strategy where each user joins with probability s. From user i's perspective, the number of other users in the multicast group will be Poisson distributed with mean $s\lambda$. Hence the unconditional expected cost share of user i will be

$$q_i(s) = \sum_{n=0}^{\infty} \frac{(s\lambda)^n \exp(-s\lambda)\bar{q}_i(n)}{n!}$$

Since the family of Poisson distributions, Poisson$(s\lambda)$, is stochastically increasing, $q_i(s)$ is decreasing in s.

Lemma 10. *1. If $q_i(0) > V$ then 0 is an NE, a pure strategy equilibrium where none of the users joins the multicast group.*
2. If $q_i(1) \leq V$ then 1 is an NE, a pure strategy equilibrium where all the users join the multicast group.
3. If $q_i(0) \geq V \geq q_i(1)$, the symmetric multi-strategy s^* such that $q_i(s^*) = V$, is the unique mixed strategy NE.

A.6.1 Information on the number of users

We restrict to *HCA* cost sharing in this section. Let U be the random variable for the cost, and $F(\cdot)$ be its distribution. f(p) being the cost of power p,

$$F(u) = G(f^{-1}(u)).$$

We assume that the base station broadcasts, N, the number of users.

Consider user i. If there are n other users in contention, the expected cost share user i, is

$$\bar{q}_i^n = \frac{\int_V^\infty (1 - F^{n+1}(u))du}{n+1}.$$

Consider the symmetric multi-strategy where each user joins with probability s. Then the unconditional expected cost share of user i will be

$$\bar{q}_i(s) = \sum_{n=0}^{N-1} \binom{N-1}{n} s^n (1-s)^{N-1-n} \bar{q}_i^n.$$

The equilibrium policy s is the solution of the equation $\bar{q}_i(s) = V$.

A.6.2 Some more information on ones own requirement

We assume a little more information: the base station tells each user i whether $f(p_i)$ is below or above V. It further broadcasts, N, the number of users having power requirements above $f^{-1}(V)$. The conditional distribution of cost for such a user

$$\tilde{F}(u) = \frac{F(u) - F(V)}{1 - F(V)}.$$

Note that for users with requirements below $f^{-1}(V)$, joining the multicast group is the dominant strategy. They also do not affect the costs of other N users. Hence we consider a noncooperative gave with N players only.

Again consider user i. If there are n other users in contention, the expected cost share user i, is

$$\tilde{q}_i^n = \frac{\int_V^\infty (1 - \tilde{F}^{n+1}(u))du}{n+1}.$$

Also consider the symmetric multi-strategy where each user joins with probability s. From user i's perspective, the number of other users in the multicast group will have Binomial $(N-1, s)$ distribution. Hence the unconditional expected cost share of user i will be

$$\tilde{q}_i(s) = \sum_{n=0}^{N-1} \binom{N-1}{n} s^n (1-s)^{N-1-n} \tilde{q}_i^n.$$

Lemma 11. *A symmetric multi-strategy s^* such that $\tilde{q}_i(s^*) = V$ is a symmetric NE.*

Remark 18. *1. As expected $s^* = 0$ is an NE.*
2. If $\tilde{q}_i^{N-1} \leq V$, $s^* = 1$ is also an NE. Thus providing less information may potentially improve the user base.

A.6.3 Properties of NEs

Theorem 24. *Assume that each user knows at least its own request. Then among all monotone coalitions achievable by some cost sharing policy, HCA achieves the smallest range.*

A.7 On the expected capacity and coverage

Consider a random realization of the network: there is a single BS, a point process that describes the location of the mobiles. All mobiles are assumed to have full knowledge. Let $\overline{S}$ be the set of all mobiles. Let there be an alternative unicast solution for any individual that costs V. Fix some cost sharing mechanism. Consider a subset $S \subset \overline{S}$ such that there is no mobile in S that can benefit from leaving the multicast group S (for the given cost sharing policy) and getting instead

V. We call such a set a "V-stable set". Assume that S is a maximal set, i.e. if we add to it another $s \in \overline{S}$ then the new set is not V-stable anymore.

We define the capacity associated with a V-stable set as the number of mobiles in S.

is this well defined? could the same problem have several different stable sets?

We define the coverage of S to be the set of locations in which if we placed another mobile, then the new set will still be V-stable.

Unless otherwize stated, coverage and capacity are defined under full information conditions.

A.7.1 Computing Capacity and Coverage

Consider a network with the BS at zero. Assume that the mobiles are located according to a stationary Poisson process with intensity λ.

HCA policy: network over the line The capacity of the system for a given value of V is a Poisson random variable with parameter $2\lambda V$. In particular, $2\lambda V$ is also the expected capacity. The coverage region is the interval $[-V, V]$.

Incremental Cost Policy: network over the plain. The capacity of the system for a given value of V is a Poisson random variable with parameter $\pi(\lambda V)^2/2$. In particular, $\pi(\lambda V)^2/2$. is also the expected capacity. The coverage region is the circle of radius V centered at zero.

The incremental cost policy over the line: Recall (A.5). This is a condition in terms of the difference of powers needed by adjacant mobiles. We find it useful to express the latter as a function of the locations of the mobiles and the BS, since coverage is understood as a geometric property. We shall use the following simple path loss attenuation model for simplicity.

If the BS transmits at power P then we have at distance d a signal of power $p = hPd^{-\alpha}$. If we have a given sensitivity threshold p that guarantees reception at reasonable quality, then the power needed to get p at a distance d is pd^α/h. Let $v(q)$ correspond to the price of transmitting at a power q.

Assume that there is a mobile at b and at $a < b$. There is a BS at the origin. Assume that the mobile at a participates in the multicast session. then that at b will participate in the multicast session if

$$V > v(pb^\alpha/h) - v(pa^\alpha/h)$$

or equivalently

$$v(pb^\alpha/h) \leq V + v(pa^\alpha/h)$$

Let v be linear. Then the condition becomes

$$b \leq \left(\frac{Vh}{p} + a^\alpha\right)^{1/\alpha} \tag{A.6}$$

Let $c := h/p$.

The number of mobiles in a multicast session: Enumerate the mobiles according to the increasing distance to the BS. Define X_0 and let X_n be the location of the nth mobile. Then the capacity N is given by

$$N = \sup\{k : (X_n)^\alpha - (X_{n-1})^\alpha < cV, \forall n = 1, ..., k\}$$

where $\sup \emptyset := 0$. The coverage is given by

$$C = \left(\frac{Vh}{p} + (X_N)^\alpha\right)^{1/\alpha}$$

Let $N(0) := N$ and define

$$N(m) = \sup\{k : (X_n)^\alpha - (X_{n-1})^\alpha < cV, \forall n = m+1, ..., k\}$$

$N(m)$ is the capacity of the system if the coverage satisfies $C > X_m$. Define $C(m)$ as

$$C(m) := \left(\frac{Vh}{p} + (X_{N(m)})^\alpha\right)^{1/\alpha}$$

A.7.2 The linear case: $\alpha = 1$

Lemma 12. *In case $\alpha = 1$ then*
(i) $N(m)$ are identically distributed, $m = 0, 1, 2, ...$
(ii) $C(m)$ are identically distributed, $m = 0, 1, 2, ...$

The expected coverage distance C can be computed as follows. Define

$$C_0 = \frac{Vh}{p}$$

(This is also the capacity reached when applying the rule that it is the furthest that pays.) The location X of the first mobile is exponentially distriubted with parameter λ. With probability $\exp(-\lambda C_0)$ we have $X > C_0$ and then $C = C_0$. With the complementary probability, $X < C_0$ so that $C > X$. In that case, $C = X + C(1)$ where $C(1)$ has the same distribution as C (due to Lemma 12). We conclude that

$$\begin{aligned} E[C] &= E[C1\{X > C_0\}] + E[(X + C)1\{X \leq C_0\}] \\ &= C_0 \exp(-\lambda C_0) + E[X + C] - E[(X + C)1\{X > C_0\}] \\ &= C_0 \exp(-\lambda C_0) + C(1 - \exp(-\lambda C_0)) + E[X] - E[(X1\{X > C_0\}] \\ &= C_0 \exp(-\lambda C_0) + E[C](1 - \exp(-\lambda C_0)) + \frac{1}{\lambda} - \exp(-\lambda C_0)(C_0 + \frac{1}{\lambda}) \end{aligned}$$

Thus the expected capacity is given by

$$E[C] = \frac{C_0 \exp(-\lambda C_0) + \frac{1}{\lambda} - \exp(-\lambda C_0)(C_0 + \frac{1}{\lambda})}{\exp(-\lambda C_0))}$$

A.7.3 $\alpha > 1$

We have:

$$b \leq \left(\frac{Vh}{p} + a^{\alpha}\right)^{1/\alpha} \leq \frac{Vh}{p} + a \tag{A.7}$$

Is it more generally, decreasing in α?
Can this be used to show that the linear case gives an upper bound?

Appendix B

Dynamics of competition: replicator dynamics

We introduce here the replicator dynamics which describes the evolution in the population of the various strategies. In the replicator dynamics, the share of a strategy in the population grows at a rate equal to the difference between the payoff of that strategy and the average payoff of the population. More precisely, consider N strategies. Let $\mathbf{x}$ be the N dimensional vector whose ith element x_i is the population share of strategy i (i.e. the fraction of the population that uses strategy i). Thus we have $\sum_i x_i = 1$ and $x_i \geq 0$. Below we denote by $J(i,k)$ the expected payoff (or the fitness) for a player using strategy i when it encounters a player with strategy k. With some abuse of notation we define $J(i,\mathbf{x}) = \sum_j J(i,j)x_j$.

B.1 Definition and properties

The replicator dynamics is defined as

$$\begin{aligned} \dot{x}_i(t) &= x_i K\Big(J(i,\mathbf{x}) - \sum_j x_j J(j,\mathbf{x})\Big) \\ &= x_i K\Big(\sum_j x_j J(i,j) - \sum_j \sum_k x_j J(j,k) x_k\Big) \end{aligned} \tag{B.1}$$

where K is a positive constant and $\dot{x}_i(t) := \mathrm{d}x_i(t)/\mathrm{d}t$. Note that the right hand side vanishes when summing over i. This is compatible with the fact that we study here the share of each strategy rather than the size of the population that uses each one of the strategies.

B.2 Replicator dynamics with delay

In Equation (B.1), the fitness of strategy i at time t has an instantaneous impact on the rate of growth of the population size that uses it. An alternative more realistic model for replicator dynamic would have some delay: the fitness acquired at time t will impact the rate of growth τ time later. We then have

$$\dot{x}_i(t) = x_i(t)K \times \tag{B.2}$$

$$\left(\sum_j x_j(t-\tau)J(i,j) - \sum_{j,k} x_j(t)J(j,k)x_k(t-\tau)\right)$$

where K is some positive constant. (It can be interpreted as a scaling factor of the fitness $J(,.,)$ or as a gain parameter that controls the speed of adaptation. More details will be given in Section

??.) The delay τ represents a time scale much slower than the physical (propagation and queueing) delays, it is related to the time scale of (i) switching from the use of one protocol to another (ii) upgrading protocols.

Bibliography

[1] J. Love, "Cell phone jammers," *methodshop.com, site createdat 09/01/2005 and updated at 04/24/2009, www.methodshop.com/gadgets/reviews/celljammers.*

[2] S. W. Shah, M. I. Babar, and M. N. Arbab, "Cell phone jammer," in *Multitopic Conference, INMIC 2008. IEEE International Issue available in Xplore*, 23-24 Dec 2008.

[3] E. Altman and L. Wynter, "Equilibrium, games, and pricing in transportation and telecommunications networks," *Networks and Spatial Economics*, vol. 4, no. 1, pp. 7–21, Mar. 2004, special issue on "Crossovers between Transportation Planning and Telecommunications".

[4] E. Altman, T. Boulogne, R. E. Azouzi, T. Jimenez, and L. Wynter, "A survey on networking games," *Computers and Operations Research*, vol. 33, no. 2, pp. 286–311, 2006.

[5] S. Lasaulce, M. Debbah, and E. Altman, "Methodologies for analyzing equilibria in wireless games," *IEEE Signal Processing Magazine*, vol. 26, no. 5, pp. 51–52, 2009, special issue on Game Theory.

[6] E. Altman and R. El-Azouzi, "La théorie des jeux non-coopératifs appliquée aux réseaux de télécommunication," *Annales des Télécommunications*, 2007, (in French).

[7] T. Bonald, "Comparison of tcp reno and tcp vegas: Efficiency and fairness," *Perform. Eval.*, pp. 307–332, 1999.

[8] T. Issariyakul and E. Hossain, *Introduction to Network Simulator NS2.* Springer, Nov. 2008, iSBN 978-0-387-71759-3.

[9] B. von Stengel, *Algorithmic Game Theory.* Cambridge: Cambridge Univ. Press, 2007, ch. Chapter 3: Equilibrium computation for two-player games in strategic and extensive form, pp. 53–78, thanks to the progressive-minded publisher, the entire book is accessible online by clicking here (username=agt1user, password=camb2agt).

[10] ——, *Handbook of Game Theory.* Amsterdam: North-Holland, 2002, vol. 3, ch. Chapter 45: Computing equilibria for two-person games, pp. 1723–1759.

[11] A. Charnes, "Constrained games and linear programming," in *Proceedings of the National Academy of Science of the U.S.A*, 1953, pp. 639–641.

[12] J. Aubin, *Optima and Equilibria, An Introduction to Nonlinear Analysis.* Springer-Verlag, Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong, Barcelona, Budapest., 1993.

[13] J. G. Proakis, *Communication Systems Engineering.* Prentice Hall International Editions, 1994.

[14] D. Barman, I. Matta, E. Altman, and R. E. Azouzi, "Tcp optimization through fec, arq and transmission power tradeoffs," in *2nd international Conference on Wired/Wireless Internet Communications WWIC*, Frankfurt (Oder), Germany, Feb. 2004.

[15] L. Galluccio, G. Morabito, and S. Pallazzo, "An analytical study of tradeoff between transmission power and fec for tcp optimization in wireless networks," in *Proceedings of IEEE INFOCOM*, 2003.

[16] P. Soldati, *Cross-Layer Optimization of Wireless Multi-hop Networks.* Royal Institute of Technology (KTH), 2007, (ISBN 978-7178-711-8).

[17] R. Cooper, *Coordination Games.* Cambridge University Press, 1998, (ISBN 0-521-57896-5).

[18] Y. Tao and Z. Wang, "Effect of time delay and evolutionarily stable strategy," *J. Theor. Biol.*, no. 187, pp. 111–116, 1997.

[19] H. Singh and S. Singh, "Energy consumption of tcp reno, new reno and sack in multi-hop wireless networks," in *ACM SIGMERTICS, Jun.*, 2002.

[20] M. Zorzi and R. Rao, "Energy efficiency of tcp in a local wireless environment," *Mobile Networks and Applications*, vol. 6, no. 3, 2001.

[21] O. Ait-Hellal, E. Altman, D. Elouadghiri, M. Erramdani, and N. Mikou, "Performance of tcp/ip: the case of two controlled sources," in *ICCC'97, Cannes, France, November 19-21*, 1997.

[22] E. Altman, R. El-Azouzi, Y. Hayel, and H. Tembine, "An evolutionary game approach for the design of congestion control protocols in wireless networks," in *Physicomnet workshop, Berlin, April 4*, 2008.

[23] "Notice of proposed rulemaking (fcc 09/93)," October 2009 – April 2010.

[24] "Consultation publique sur la neutralité du net, du 9 avril au 17 mai 2010."

[25] N. Kroes, "vice president of the european commission commissioner for the digital agenda," *Press release RAPID*, April 2010.

[26] R. Hahn and S. Wallsten, "The economics of net neutrality," *Economists' Voice*, vol. 3, no. 6, pp. 1–7, 2006.

[27] K. Cheng, S. Bandyopadhyay, and H. Gon, "The debate on net neutrality: A policy perspective," *Information Systems Research*, June 2008.

[28] R. Ma, D.-M. Chiu, J. Lui, V. Misra, and D. Rubenstein, "Interconnecting eyeballs to content: A shapley value perspective on isp peering and settlement," in *Proc. Int'l Workshop on Economics of Networked Systems (NetEcon)*, 2008, pp. 61–66.

[29] ——, "On cooperative settlement between content, transit and eyeball internet service providers," in *Proc. ACM Int'l Conference on Emerging Networking EXperiments and Technologies (CoNEXT)*, 2008.

[30] P. Hande, M. Chiang, R. Calderbank, and S. Rangan, "Network pricing and rate allocation with content provider participation," in *Proc. IEEE INFOCOM*, 2009.

[31] J. Musacchio, G. Schwartz, and J. Walrand, "A two-sided market analysis of provider investment incentives with an application to the net-neutrality issue," *Review of Network Economics*, vol. 8, no. 1, 2009.

[32] H. Kameda, J. Li, C. Kim, and Y. Zhang, *Optimal Load Balancing in Distributed Computer Systems.* Springer-Verlag, London, Telecommunications Networks and Computer Systems (TNCS) Series, 1997.

[33] J. G. Wardrop, "Some theoretical aspects of road traffic research communication networks," *Proc. Inst. Civ. Eng., Part 2*, vol. 1, pp. 325–378, 1952.

[34] A. Haurie and P. Marcotte, "On the relationship between nash-cournot and wardrop equilibria," *Networks*, vol. 15, pp. 295–308, 1985.

[35] J. Weibull, *Evolutionary Game Theory*. M.I.T. Press, 1995.

[36] D. D. Yao, "S-modular games, with queuing applications," *Queuing Systems: Theory and Applications*, vol. 21, pp. 449–475, 1995.

[37] D. Topkis, "Equilibrium points in nonzero-sum n-person submodular games," *SIAM J. Control and Optimization*, vol. 17, pp. 773–787, Nov. 1979.

[38] A. Orda, R. Rom, and N. Shimkin, "Competitive routing in multi-user communication networks," *IEEE/ACM Transactions on Networking*, vol. 1, no. 5, pp. 510–520, 1993.

[39] P. Marbach, "Priority service and max-min fairness," *IEEE/ACM Transactions on Networking (TON)*, vol. 11, pp. 733–746, October 2003.

[40] A. Bovopoulos and A. Lazar, "Optimal routing and flow control of a network of parallel procesors with individual buffers," *Proceedings of the 23rd Allerton Conference on Communication, Control, and Computing*, pp. 564–573, 1985.

[41] M. T. Hsiao and A. A. Lazar, "Optimal decentralized flow control of Markovian queueing networks with multiple controllers," *Performance Evaluation*, vol. 13, pp. 181–204, 1991.

[42] F. P. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications, special issue on Elastic Services over ATM networks*, vol. 8, no. 1, 1997.

[43] F. Kelly, A. Maulloo, and D. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, pp. 237–252, 1998.

[44] S. H. Low and D. E. Lapsley, "Optimization flow control–I: basic algorithm and convergence," *IEEE/ACM Transactions on Networking*, vol. 7, no. 6, pp. 861–874, 1999.

[45] K. Park, M. Sitharam, and S. Chen, "Quality of service provision in noncooperative networks with diverse user requirements," *Decision Support Systems, Special Issue on Information and Computation Economies*, vol. 28, pp. 101–122, 2000.

[46] J. Rosen, "Existence and uniqueness of equilibrium points for concave n-person games," *Econometrica*, vol. 33, pp. 520–534, 1965.

[47] A. Legrand and C. Touati, "Non-cooperative scheduling of multiple bag-of-task appplications," in *Proceedings of the 25th Conference on Computer Communications (INFOCOM'07)*, Alaska, USA, May 2007.

[48] M. Barnsley, *Fractals Everywhere*. Morgan Kaufmann, 1988.

[49] M. Detyniecki, "Mathematical aggregation operators and their application to video querying," Ph.D. dissertation, University of Paris 6, France, 2000.

[50] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Transactions on Networking*, vol. 8, no. 5, pp. 556–567, 2000.

[51] R. K. Jain, dah Ming W. Chiu, and W. R. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," DEC Research Report, Tech. Rep. DEC-TR-301, September 1984. [Online]. Available: http://www.rajjain.com/papers/fairness.htm

[52] C. Stein and J. Wein, "On the existence of schedules that are near-optimal for both makespan and total weighted completion time," *Operational research letters*, vol. 21, no. 3, pp. 115–122, 1997.

[53] E. Koutsoupias and C. Papadimitriou, "Worst-case equilibria," in *STACS*, 1998.

[54] C. H. Papadimitriou and M. Yannakakis, "On the approximability of trade-offs and optimal access of web sources," in *FOCS '00: Proceedings of the 41st Annual Symposium on Foundations of Computer Science*. Washington, DC, USA: IEEE Computer Society, 2000, p. 86.

[55] C. Touati, E. Altman, and J. Galtier, "Generalised Nash bargaining solution for banwidth allocation," *Computer Networks*, 2006 (to appear).

[56] H. Kameda, "Bounds on benefits and harms of adding connections to noncooperative networks," in *NETWORKING 2004*, ser. LNCS, N. Mitrou, K. Kontovasilis, G. N. Rouskas, I. Iliadis, and L. Merakos, Eds., vol. 3042. Springer Verlag, 2006, pp. 405–417.

[57] A. Glazer and R. Hassin, "?/m/1: On the equilibrium distribution of customer arrivals," *European J. of OR*, vol. 13, pp. 146–150, 1983.

[58] A. Elcan, "Optimal customer return rate for an m/m/1 queueing system with retrials," *Probability in the Engineering and Informational Sciences*, vol. 8, pp. 521–539, 1994.

[59] R. Hassin and M. Haviv, "On optimal and equilibrium retrial rates in a busy system," *Prob. in the Eng. and Informational Sciences*, vol. 10, pp. 223–227, 1996.

[60] V. G. Kulkarni, "On queueing systems with retrials," *Journal of Applied Probability*, vol. 20, pp. 280–289, 1983.

[61] Y. Jin and G. Kesidis, "Equilibria of a noncooperative game for heterogeneous users of an aloha network," *IEEE Comm. Letters*, vol. 6, no. 7, pp. 282–284, 2002.

[62] A. B. MacKenzie and S. B. Wicker, "Selfish users in aloha: A game theoretic approach," in *Proceedings of the Fall 2001 IEEE Vehicular Technology Conference*, 2001.

[63] D. Braess, "Uber ein paradoxen der werkehrsplannung," *Unternehmenforschung*, vol. 12, pp. 256–268, 1968.

[64] B. Calvert, W. Solomon, and I. Ziedins, "Braess's paradox in a queueing network with state-dependent routing," *J. Appl. Prob.*, vol. 34, pp. 134–154, 1997.

[65] J. E. Cohen and C. Jeffries, "Congestion resulting from increased capacity in single-server queueing networks," *IEEE/ACM Trans. on Networking*, vol. 5, no. 2, pp. 1220–1225, April 1997.

[66] J. E. Cohen and F. P. Kelly, "A paradox of congestion in a queuing network," *J. Appl. Prob.*, vol. 27, pp. 730–734, 1990.

[67] Y. A. Korilis, A. A. Lazar, and A. Orda, "Avoiding the braess paradox in non-cooperative networks," *Journal of Applied Probability*, vol. 36, pp. 211–222, 1999.

[68] R. J. Aumann, "Subjectivity and correlation in randomized strategies," *Journal of Mathematical Economics*, 1974.

[69] R. Hassin, "On the advantage of being the first server," *Management Science*, vol. 42, pp. 618–623, 1996.

[70] E. Altman, T. Jiménez, R. Nunez-Queija, and U. Yechiali, "Queueing analysis for optimal routing with partial information," manuscript, 2001.

[71] I. Adiri and U. Yechiali, "Optimal priority purchasing and pricing decisions in nonmonopoly and monopoly queues," *Operations Research*, 1974.

[72] R. Hassin and M. Haviv, "Equilibrium threshold strategies: the case of queues with priorities," *Operations Research*, vol. 45, pp. 966–973, 1997.

[73] B. Peleg and P. Sudhoelter, "Introduction to the theory of cooperative games, second edition, 123, theory and decision library," G. E. W. L. (Vienna), M. P. G. Eberlein (Munich), Series C: Game Theory, and O. Research, Eds., vol. 34, 2nd Edition. Springer-Verlag Berlin Heidelberg, 2007.

[74] S. C. Littlechild and G. Owen, "A simple expression for the shapley value in a special case," *Management Science*, vol. 20, no. 3, pp. 370–372, Nov. 1973.

[75] S. C. Littlechild, "A simple expression for the nucleolus in a special case," *International Journal of Game Theory*, vol. 3, no. 1, pp. 21–29, March 1974.

[76] W. Thomson, "Cost allocation and airport problems," *Working paper, No. 538, Rochester-Center for Economic Research*, Dec. 2007.

[77] R. B. Myerson, "Large poisson games." *Journal of Economic Theory*, vol. 94, pp. 1–45, 2000.

[78] P. Penna and C. Ventre, "Sharing the cost of multicast transmissions in wireless networks," in *Proceedings of SIROCCO 2004, LNCSl, 3104*, 2004, pp. 255–266.

[79] V. Bilo, M. Flammini, G. Melideo, L. Moscardelli, and A. Navarra, "Sharing the cost of multicast transmissions in wireless networks," *Theoretical Computer Science*, vol. 369, pp. 269–284, 2006.

[80] V. Bilo, M. Flammini, G. Melideo, and L. Moscardelli, "On nash equilibria for multicast transmissions in ad-hoc wireless networks," *Wireless Networks*, vol. 14, pp. 147–157, 2008.

[81] S. H. Tijs and T. S. H. Driessen, "Game theory and cost allocation problems," *Management Science*, vol. 32, no. 8, pp. 1015–1028, August 1986.

[82] L. S. Shapley, "Cores of convex games," *International Journal of Game Theory*, vol. 1, pp. 11–26, Dec. 1971.

[83] B. Dutta and D. Ray, "A concept of egalitarianism under participation constraints," *Econometrica*, vol. 57, no. 3, pp. 615–635, May 1989.

[84] K. Jain and V. V. Vazirani, "Equitable cost allocations via primal-dual-type algorithms," in *34th annual ACM symposium on Theory of computing*, 2002, pp. 313–321.

[85] C.-H. Yeh, "An alternative characterization of the nucleolus in airport problems, mimeo, institute of economics, academia sinica, taipei, taiwan."

[86] H. Moulin, "Incremental cost sharing: Characterization by coalition strategy-proofness," *Social Choice and Welfare*, vol. 16, no. 2, pp. 279–320, 1999.

[87] B. Dutta, "The egalitarian solution and reduced game properties in convex games," *International Journal of Game Theory*, vol. 19, no. 2, pp. 153–169, 1990.

[88] T. Sonmez, "Population monotonicity of the nucleolus on a class of public good problems, mimeo, rochester-center for economic research," 1994.