NET Institute*

www.NETinst.org

Working Paper #10-12

September 2010

Mobile Termination and Consumer Expectations under the Receiver-Pays Regime

Sjaak Hurkens
CSIC and IESE Business School, University of Navarra

Ángel L. López
IESE Business School, University of Navarra

* The Networks, Electronic Commerce, and Telecommunications (“NET”) Institute, http://www.NETinst.org, is a non-profit institution devoted to research on network industries, electronic commerce, telecommunications, the Internet, “virtual networks” comprised of computers that share the same technical standard or operating system, and on network issues in general.
Mobile Termination and Consumer Expectations under the Receiver-Pays Regime

Sjaak Hurkens and Ángel L. López*

September 2010

Abstract

We analyze how termination charges affect retail prices when taking into account that receivers derive some utility from a call and when firms may charge consumers for receiving calls. A novel feature of our paper is that we consider passive self-fulfilling expectations and do not allow for negative reception charges. We reconfirm the finding of profit neutrality when firms cannot use termination-based price discrimination and show that connectivity is prone to breakdown.

Keywords: Bill and Keep; Call externality; Access Pricing; Interconnection; Receiver pays; Consumer Expectations

JEL classification: D43; K23; L51; L96

1 Introduction

Although the telecommunications sector has been liberalized in most industrialized countries, some regulation remains. A clear example is call termination on mobile telephone networks. Mobile operators interconnect their networks so that their customers can communicate with the customers of other networks. This requires mobile operators to provide a wholesale service called ‘call termination’, whereby each completes a call made to one of its subscribers by a caller on another network. In most countries, call termination is provided in exchange for a fee or access charge, which is also called mobile termination rate. Indeed, above-cost termination rates is a notorious feature of most of the European markets. A second feature

*Hurkens: Institute for Economic Analysis (CSIC) and Public-Private Sector Research Center (IESE Business School, University of Navarra, Spain), sjaak.hurkens@iae.csic.es. López: Public-Private Sector Research Center, IESE Business School, University of Navarra, Spain, alopezr@iese.edu. Financial support from the Net Institute, http://www.Netinst.org is gratefully acknowledged.
of these markets is that users do not pay for receiving calls, i.e., the so-called "calling party
pays" (CPP) principle.

There are two types of call termination: termination of calls originated on the fixed-line
telephone network (fixed-to-mobile FTM termination) and termination of calls originated on
other mobile networks (mobile-to-mobile MTM termination). There is a consensus that if
FTM termination rates are left unregulated, then mobile operators will unilaterally set too
high termination rates. As a mobile user usually join just one mobile network, the network
to which she subscribes is the only that can provide the service of call termination. Unlike
FTM rates, MTM rates inflate the cost of the off-net calls (i.e., those calls originated on a
network and completed on a different network), and thus affect competition (and efficiency)
in the retail market for mobile telephony.

Starting with the seminal works of Armstrong (1998) and Laffont, Rey and Tirole (1998a,
b) (henceforth ALRT), a burgeoning literature that analyses the impact of termination rates
on competition has emerged.\(^1\) Nevertheless, most of the theoretical work on mobile network
interconnection typically assumed that consumers derive utility only from making calls,
ignoring the existence of call externalities — that is, the fact that not only callers but also
receivers of a call enjoy a positive benefit.\(^2\)

The possibility that the receiving party enjoy benefits from a call is clearly important
for the manner in which firms compete in the retail market. Once it is recognized that
consumers enjoy benefits from receiving a call, it follows that they are prepared to pay for
this. Indeed, in some countries (e.g. Canada, Singapore, Hong Kong and the United States)
mobile operators charge their subscribers for the calls they receive. The objective of this
paper is to explore the implications of call externalities and termination rates for pricing
under the possibility that mobile operators can charge both outgoing and incoming calls,
which is known as the receiver-pays regime.

An incipient literature has started to examine the relationship between termination rates
and equilibrium prices in an environment with call externalities (Kim and Lim [2001]; De-
Graba [2003]; Hahn [2003]; Laffont et al. [2003]; Jeon et al. [2004]; Berger [2004, 2005];

---

\(^1\) For a complete review of the literature on access charges see Armstrong (2002), Vogelsang (2003) and
Peitz et al. (2004).

\(^2\) One assumption that is invoked to justify the absence of call externalities in models of network
competition is that call externalities could be largely internalized by the parties (see Competition Commission
[2003, paras 8.257 to 8.260]). However, as argued by Hermalin and Katz (2004, p. 424), "this assumption is
applicable only to a limited set of situations in which either the communicating parties behave altruistically
or have a repeated relationship". Additionally, Harbord and Pagnozzi (2010) argue that the empirical basis
for the internalization of call externalities is unclear.

\(^3\) Jeon et al. (2004) provides a short overview of the literature on competition in the presence of call
externalities.
et al. (2003; LMRT hereafter), Jeon et al. (2004; JLT hereafter), Hermadin and Katz (2006), Cambini and Valletti (2008) and López (2010) are the papers closest to ours. LMRT analyze Internet backbone competition and assume that there exist two types of users: websites (senders) and consumers (receivers). Hermadin and Katz study whether termination charges can induce carriers to internalize the externalities that arise when both senders and receivers of telecommunications messages enjoy benefits. But in contrast to the framework of LMRT, in which there are two different types of users, they consider that any given user has a one-half chance of being a sender and a one-half chance of being a receiver. In JLT, López (2010) and this paper, however, every consumer both sends and receives traffic, and moreover obtains surplus from and is charged for placing and receiving calls. One common feature of these papers is that they study the impact of termination charges on equilibrium calling and reception charges in a framework where outgoing and incoming calls are unrelated. However, Cambini and Valletti (2008) argue that an exchange of information may yield further exchanges (for example when calls made and received are complements), so they consider a framework in which incoming and outgoing calls are interdependent. In the present paper we characterize the equilibrium when two interconnected networks charge both for outgoing and incoming calls, and the calls made and received are independent of each other. We develop further the analysis of JLT and obtain new results that have implications for retail pricing. However, the novelty of our analysis lies in studying how consumer expectations affect equilibrium end-user prices (and so equilibrium profit and welfare).

As it is well known, consumers expectations are crucial whenever externalities exist. Existing models of network competition in the spirit of ALRT implicitly assume that consumers have what we term ‘rationally responsive expectations’. By this we mean that first firms set prices, second consumers form expectations about network sizes depending on the prices charged by the firms, and third consumers make optimal subscription decisions (given the prices and their expectations). This assumption implies that any change of a price by one firm is assumed to lead to an instantaneous rational change in expectations of all consumers, such that, given these new expectations, optimal subscription decisions will lead realized and expected network sizes to coincide.

In Hurkens and López (2010) we show that the way consumers form expectations about network sizes is crucial for the relationship between termination charges and equilibrium profit. This is the case in which networks compete in nonlinear pricing, and under network-based (i.e., on-net/off-net) price discrimination and the CPP principle. Implicitly assuming rationally responsive expectations, Laflont et al. (1998b) show that profit is strictly decreasing in termination charge. Building on their analysis, Gans and King (2001) show furthermore that firms strictly prefer below cost termination charges. Intuitively, if termi-
nation charge is above cost, then off-net calls will be more expensive than on-net calls. As there is a price differential between on- and off-net calls, consumers care about the size of each network (the so-called ‘tariff-mediated network externalities’) and so they prefer to join the larger network. Consequently, acquisition costs are reduced, which in turn intensifies competition for subscribers and results in lower subscription fees. As a matter of fact this result is at odds with real world observations since regulators around the world, and especially in the European Union, are concerned about too high termination charges. But at the same time this result has been shown to be very robust. For example, it holds for any number of networks [Calzada and Valletti, 2008], in the presence of call externalities [Berger, 2005], and when networks are asymmetric [López and Rey, 2009]. Also, Hurkens and Jeon (2009) show that this result holds when there are both network externalities (i.e., elastic subscription demand as in Dessein, 2003) and network-based price discrimination.

Nevertheless, in Hurkens and López (2010) we observe that a seemingly innocuous twist of the modeling of consumer expectations reconcile the puzzle: firms prefer termination charges above cost, and socially optimal termination charges are below or at cost (depending on the case that is under consideration). In particular, we relax the assumption of rationally responsive expectations and replace it by one of fulfilled equilibrium expectations (also termed passive -self-fulfilled- expectations), which implies the following: first consumers form expectations about network sizes, then firms compete, and finally consumers make optimal subscription or purchasing decisions, given the expectations and the prices, so that in equilibrium realized and expected network sizes coincide.

It is worth mentioning that a few recent papers also attempt to reconcile the mentioned puzzle. Armstrong and Wright (2009), Jullien, Rey and Sand-Zantman (2010), and Hoernig, Inderst and Valletti (2009) have in common that they introduce additional realistic features of the telecommunication industry into the Laffont, Rey and Tirole (1998b) framework. They show that for some parameter range (and under rationally responsive expectations) joint profits increase as the termination charge increases above the cost. However,

---

4This result is robust to the inclusion of call externalities, an arbitrary number of mobile operators, asymmetric networks and elastic subscription demand.

5This concept was first proposed by Katz and Shapiro (1985).

6Armstrong and Wright (2009) argue that if MTM and FTM termination charges must be chosen uniformly, as is in fact the case in most European countries, firms will trade off desirable high FTM and desirable low MTM charges and arrive at some intermediate level, which may well be above cost.

7Jullien, Rey and Sand-Zantman (2010) argue that the willingness to pay for subscription is related to the volume of calls. They introduce two types of users in the framework of ALRT: light users and heavy users. Light users only receive calls and are assumed to have an elastic subscription demand. Instead, full participation is assumed for heavy users, who can place calls and obtain a fixed utility from receiving calls.

8Hoernig, Inderst and Valletti (2009) consider the existence of calling clubs so that the calling pattern is not uniform but skewed.
contrary to Hurkens and López (2010) these papers conclude that the need to regulate termination charges is reduced because the socially optimal termination charge would also be above cost.

The implications of the assumption of how consumers form expectations on the relationship between termination charges and equilibrium profit under the CPP principle motives us to revisit the analysis of network competition in the presence of call externalities and the receiver-pays regime by considering passive (self-fulfilled) expectations. A second motivation for our analysis is provided by the current practice in the European Union that basically consists of progressive reductions in termination charges. Until recently, most regulatory authorities in Europe set termination charges above the (marginal) cost of termination so as to recover the fixed and common costs of an hypothetical efficient network operator incurred in providing services in the retail and wholesale markets. In May 2009, the European Commission recommended national regulatory authorities to set termination rates based on the costs (i.e., the actual incremental cost of providing call termination — without allowing for common costs) incurred by an efficient operator.\(^9\) The European Commission’s view was also supported by the European regulators group, who in the Common Position adopted on February 2008\(^10\) decided to take a position in favor of setting a unique and uniform termination rate for all network operators at the cost incurred by an hypothetical efficient operator. As a result, the average MTR in Europe could drop from about 8.55 euro cents per minute at the end of 2009, to approximately 2.5 euro cents per minute by 2012 [see Harbord and Pagnozzi, 2010]. In light of these announcements, some large European mobile operators as for instance Vodafone, warned the European Union that cutting termination rates could mean the end of handset subsidies for consumers and lead to a price increase. Furthermore, Vodafone claimed that cutting termination rates could result in a US style business model, where users pay for both placing and receiving calls.

This issue has been explored (under rationally responsive expectations) in JLT (2004) and more recently in López (2010)\(^11\). On the one hand, in the absence of network-based price discrimination, mobile operators charge calls and call receptions at their off-net cost (even if they are asymmetric in terms of market shares). This is the so-called ‘off-net-cost pricing

---


\(^11\)López (2010) generalizes the framework of JLT by allowing a random noise in both the callers’ and receivers’ utilities, by removing the assumption (at some stages) the assumption of a given proportionality between the utility functions, and by allowing asymmetry between mobile operators with respect to the number of locked-in customers.
principle. Hence operators charge incoming calls only when the termination charge is below cost (so as to recover the cost of providing the service of call termination). This is in line with the fact that the countries where mobile operators apply the receiver-pays regime, cited above, typically present below-cost termination charges. On the other hand, when mobile operators can differentiate their calling and reception charges according to whether the communication is on- or off-net, connectivity is prone to break down. The reason is that off-net calling and reception charges allow network operators to create direct externalities on the customers of rival operators. If, for example, the callers obtain more utility than the receivers from a given call, the attractiveness of the offer of the network where the call is received will be reduced in comparison with the rival’s offer. Therefore, to avoid a lose of attractiveness, the terminating network will break connectivity by charging a too high reception price.

Our paper examines the determination of competitive retail and reception charges in the presence of passive (self-fulfilled) expectations. In particular, we explore the relationship between termination charges and retail prices in the presence and absence of network-based price discrimination. After introducing the setting, we examine the case of no network-based price discrimination. We observe that the off-net-cost pricing principle is robust to the way consumers form expectations about network sizes. The reason is that mobile operators set marginal prices at the opportunity cost of ‘stealing’ the customers away from the rival operators (this maximizes consumer surplus, which can then be extracted through the fixed fee). As marginal prices do not depend on market shares, consumer expectations do not alter them. This result has two implications. First, in equilibrium profit is neutral to the level of the termination charge. Second, mobile operators only charge for incoming calls when the termination charge is below cost, as it is in the presence of rationally responsive expectations. By the same token, one would conclude that reception charges are negative (firms subsidize call receptions) when the termination charge is above cost. Hence we discuss the case in which the termination charge is above cost but reception charges cannot be negative. The analysis concludes by determining the socially optimal prices: As optimality requires a positive reception charge, optimal termination charge must be strictly below cost.

We then consider the case of network-based price discrimination. We show that connectivity is prone to breakdown. The only possibility for a symmetric equilibrium without connectivity breakdown exists when termination mark-up is negative. In this case the equilibrium is characterized by zero reception charge for off-net calls and positive reception charges.

---

12 The off-net-cost pricing principle dates back to LMRT, who found this pricing rule in a framework for Internet backbone competition.
13 Cambini and Valletti (2008) obtain the same result in their framework of information exchange between calling parties with interdependency among outgoing and incoming calls.
for on-net calls. The reason for our result to strongly differ from the ones of JLT (2004), who
describe equilibrium candidates for not too negative termination mark-ups is not caused by
our assumption about expectations. The characterization of usage prices is done in both
cases by assuming market shares constant (by adjusting the fixed fee accordingly). Hence,
we find the same candidates. The expectations play a role in determining the equilibrium
fixed fees and determining equilibrium profits. However, we point out the local convexity
of the profit function with respect to off-net reception charge. Since we bound reception
charges below by zero, we sometimes find a candidate solution with zero off-net reception
charges. If JLT (2004) would have discovered the non-concavity of the profit function, their
conclusion should have been that any symmetric equilibrium has connectivity breakdown.

The paper is organized as follows. In Section 2 we describe the model. Section 3 examines
competition when network operators cannot discriminate on the basis of where the call
terminates. Section 4 allows networks to set different prices for on-net and off-net calls.
Section 5 concludes. The technical appendix presents some useful derivatives needed to
follow the analysis of the paper.

2 The model

We consider the framework developed by JLT (2004), which extends the traditional frame-
work of network competition by allowing receivers to obtain utility from receiving calls and
firms to charge call receptions.

There are two network operators, \(i = 1, 2\), each providing full coverage.

Cost structure. The fixed cost to serve each subscriber is \(f\), whereas \(c_O\) and \(c_T\) denote
the marginal cost of providing a telephone call borne by the originating and terminating
networks. The marginal cost of an on-net call is then \(c = c_O + c_T\). Network operators pay
each other a reciprocal access charge \(a\) when a call initiated on a network is terminated on
a different network.\(^{14}\) The termination mark-up is equal to:

\[
m = a - c_T.
\]

The perceived cost of calls is the true cost \(c\) for on-net calls, augmented by the termination
mark-up for the off-net calls \(c_O + a = c + m\) for the caller’s network. The marginal cost of
an off-net call is \(c_T - a = -m\) for the receiver’s network.

Retail pricing. We consider competition in nonlinear pricing and two different cases:
i) Competition in the absence of network-based (i.e., on-net/off-net) price discrimination

\(^{14}\)Reciprocity means that a network pays as much for termination of a call on the rival network as it
receives for completing a call originated on the rival network.
(Section 3): Network $i$ offers three-part tariffs $\{F_i, p_i, r_i\}$, where $F_i$ is the monthly subscriber charge, $p_i$ is the per-unit calling price and $r_i$ is the per-unit reception charge; ii) Competition under network-based discrimination (Section 4): Network $i$ offers five-part tariffs of the form: $\{F_i, p_i, \tilde{p}_i, r_i, \tilde{r}_i\}$, where $\tilde{p}_i$ and $\tilde{r}_i$ denote the off-net calling and reception charges.

**Market shares.** The networks (i.e., firms) sell a differentiated but substitutable product. Consumers are uniformly distributed on the segment $[0, 1]$ and the two networks are located at the two extremities of the segment ($x_1 = 0, x_2 = 1$). Given income $y$, a consumer located at $x$ and joining network $i$ has utility

$$y + v_0 - t|x - x_i| + w_i,$$

where $v_0$ represents a fixed surplus from being connected to either network (it is assumed to be large enough so that all consumers want to subscribe to one network), $t|x - x_i|$ is the cost of subscribing to a network with "address" $x_i$, and $w_i$ is the net surplus of a network-$i$ consumer from making and receiving calls on that network. Network 1’s market share is given by

$$\alpha_1 = \frac{1}{2} + \sigma(w_1 - w_2),$$

where $\sigma \equiv 1/2t$ measures the degree of substitutability between the two networks. As there is full participation, 2’s market share is $\alpha_2 = 1 - \alpha_1$.

**Individual demand.** Subscribers obtain positive utility from making and receiving calls. The caller’s utility from making a call of length $q$ minutes is $u(q)$, whereas the receiver’s is $\tilde{u}(q)$ from receiving a call of that length. $u(\cdot)$ and $\tilde{u}(\cdot)$ are twice continuously differentiable, and concave. For tractability, we assume that

$$\tilde{u}(q) = \beta u(q) \quad \text{with } \beta > 0.$$  

We consider the case in which callers and receivers can hang up. To avoid multiplicity of equilibria,\textsuperscript{15} we assume that the utility that a receiver derives from receiving a call is subject to a noise $\varepsilon$: 

$$\tilde{u}(q) + \varepsilon q.$$  

$\varepsilon$ follows the distribution function $F(\cdot)$, with wide enough support $[\underline{\varepsilon}, \overline{\varepsilon}]$, zero mean, and density function $f(\cdot)$, which is strictly positive for all $\varepsilon$ in the support. Additionally, $\varepsilon$ is identically and independently distributed for each caller-receiver pair.

\textsuperscript{15}In the absence of noise and assuming that the caller determines the volume of calls, we have that from the viewpoint of networks and subscribers only the sum $\{F_i + r_iq\}$ matters, not its composition. As a result, different combinations of $F_i$ and $r_i$ are feasible equilibria but nonequivalent since each combination may affect differently the rival network.
As receivers are allowed to hang up, for a given pair of prices \((p_i, r_j)\) the length of a call from a caller of network \(i\) to a receiver of network \(j\) is given by \(q(\max(p_i, (r_j - \varepsilon)/\beta))\). Therefore, the volume of calls from network \(i\) to network \(j\) is \(\alpha_i \alpha_j D(p_i, r_j)\) with

\[
D(p_i, r_j) \equiv [1 - F(r_j - \beta p_i)]q(p_i) + \int_{\varepsilon}^{r_j - \beta p_i} q \left( \frac{r_j - \varepsilon}{\beta} \right) f(\varepsilon) d\varepsilon.
\]

Similarly, the utility that a network-\(i\) consumer obtains from placing calls to network-\(j\) consumers is \(\alpha_j U(p_i, r_j)\) with

\[
U(p_i, r_j) \equiv [1 - F(r_j - \beta p_i)]u(q(p_i)) + \int_{\varepsilon}^{r_j - \beta p_i} u \left( q \left( \frac{r_j - \varepsilon}{\beta} \right) \right) f(\varepsilon) d\varepsilon.
\]

Notice that

\[
\frac{\partial U(p_i, r_j)}{\partial p_i} = p_i \frac{\partial D(p_i, r_j)}{\partial p_i}.
\]

The utility that a network-\(j\) consumer obtains from receiving calls from network-\(i\) consumers is \(\alpha_i \tilde{U}(p_i, r_j)\) with

\[
\tilde{U}(p_i, r_j) \equiv \int_{r_j - \beta p_i}^{r_j} \left[ \tilde{u}(q(p_i)) + \varepsilon q(p_i) \right] f(\varepsilon) d\varepsilon
\]

\[
+ \int_{\varepsilon}^{r_j - \beta p_i} \left[ \tilde{u} \left( q \left( \frac{r_j - \varepsilon}{\beta} \right) \right) + \varepsilon q \left( \frac{r_j - \varepsilon}{\beta} \right) \right] f(\varepsilon) d\varepsilon.
\]

And,

\[
\frac{\partial \tilde{U}(p_i, r_j)}{\partial r_j} = r_j \frac{\partial D(p_i, r_i)}{\partial r_j}.
\]

We make the standard assumption of a balanced calling pattern, which means that the percentage of calls originating on a given network and completed on another given (including the same) network is equal to the fraction of consumers subscribing to the terminating network.\(^{16}\)

**Timing.** We assume that the terms of interconnection are negotiated or established by a regulator first. Then, for a given access charge \(a\) (or equivalently, a given \(m\)) the timing of the game is the following:

1. Consumers form expectations about the number of subscribers of each network \(i\) \((\beta_i)\) with \(\beta_1 \geq 0, \beta_2 \geq 0\) and \(\beta_1 + \beta_2 = 1\).

\(^{16}\)Dessein (2003, 2004) examines how unbalanced calling patterns between different customer types affect retail competition when network operators compete in the presence of the caller-pays regime.
2. Firms take these expectations as given and choose simultaneously retail tariffs: i) in the absence of network-based price discrimination: $T_i = (F_i, p_i, r_i)$ for $i = 1, 2$; ii) in the presence of network-based price discrimination: $T_i = (F_i, p_i, \hat{p}_i, r_i, \hat{r}_i)$ for $i = 1, 2$.

3. Consumers make rational subscription and consumption decisions, given their expectations and given the networks’ tariffs.

Therefore, market share $\alpha_i$ is a function of prices and consumer expectations. Self-fulfilling expectations imply that at equilibrium $\beta_i = \alpha_i$.

3 No Network-Based Price Discrimination

In this section, we consider (for a given reciprocal access charge $a$ and consumer expectations $\beta_1$ and $\beta_2$) competition in the presence of the receiver-pays regime and under the assumption of nondiscriminatory pricing (network operators are not allowed to charge their customers different prices for calls terminating on- and off-net).

Given the balanced calling pattern assumption and consumer expectations $\beta_1$ and $\beta_2$, the surplus from subscribing to network $i$ (gross of transportation costs) is given by

$$w_i = \phi_i(\beta_i, p_i, r_i, p_j, r_j) - F_i$$

with

$$\phi_i(\beta_i, p_i, r_i, p_j, r_j) = \beta_i U(p_i, r_i) + \beta_j U(p_i, r_j) + \beta_i \tilde{U}(p_i, r_i) + \beta_j \tilde{U}(p_j, r_i)$$

$$-p_i [\beta_i D(p_i, r_i) + \beta_j D(p_i, r_j)] - r_i [\beta_i D(p_i, r_i) + \beta_j D(p_j, r_i)].$$

When consumers’ expectations are assumed passive we have that $w_i$ is a function of expectations and prices, instead of market shares and prices as it is in the case of rationally responsive expectations. The passive expectations assumption simplifies the analysis and, as we will show below, does not change the results in the absence of price discrimination.

The profit of network $i$ can be written as (for $i \neq j = 1, 2$):

$$\pi_i = \alpha_i [\alpha_i (p_i - c) D(p_i, r_i) + \alpha_j (p_i - c - m) D(p_i, r_j) + \alpha_j m D(p_j, r_i)$$

$$+ r_i (\alpha_i D(p_i, r_i) + \alpha_j D(p_j, r_i)) + F_i - f].$$

Adjusting $F_i$ so as to maintain net surpluses $w_1$ and $w_2$ and thus market shares constant,
leads network $i$ to set $p_i$ and $r_i$ so as to maximize

$$
\pi_i = \alpha_i [\alpha_i (p_i - c) D(p_i, r_i) + \alpha_j (p_i - c - m) D(p_i, r_j) + \alpha_j m D(p_j, r_i) \\
+ r_i (\alpha_i D(p_i, r_i) + \alpha_j D(p_j, r_i)) + \phi_i (\beta_i, p_i, r_i, p_j, r_j) - \phi_j (\beta_j, p_j, r_j, p_i, r_i) \\
+ F_j - \frac{1}{\sigma} (\alpha_i - \frac{1}{2}) - f],
$$

Assume that $r_i = r_j = r$, by differentiating (5) with respect to $p_i$ and using (2), we obtain the following first-order condition:

$$
p_i = c + \alpha_j m - \alpha_i r. \quad (6)
$$

Similarly, assuming $p_i = p_j = p$, and by differentiating (5) with respect to $r_i$ and using (3), we obtain the following first-order condition:

$$
r_i = \alpha_i c - \alpha_j m - \alpha_i p. \quad (7)
$$

If $p_i = p$ and $r_i = r$, equations (6) and (7) simplify to

$$
p = c + m, \quad (8)
$$

$$
r = -m. \quad (9)
$$

Thus, in equilibrium $p$ and $r$ do not depend on market shares, and network operators charge calls and call receptions at their off-net cost.\textsuperscript{17} This is the so-called ‘off-net-cost pricing principle’\textsuperscript{18}: Each network sets prices for a subscriber’s outgoing and incoming traffic at the marginal cost that it would incur if all other subscribers belonged to the rival network. To understand this result, notice that the off-net cost is also the opportunity cost of stealing the customers away from the rival network.\textsuperscript{19} As usual with two-part tariffs, firms set the marginal price(s) at marginal cost so as to maximize the consumer surplus, which can then be extracted through the fixed part. JLT and López (2010) also find this pricing rule under the assumption of rationally responsive expectations. Therefore the off-net-cost pricing principle is robust to the assumption of consumer expectations. The reason is that

\textsuperscript{17}This equilibrium is the unique symmetric equilibrium and exists for $m \geq -\left(\frac{\beta}{1+\beta}\right) c$ (see proof of Proposition 6 in Jeon et al. [2001, Appendix 3]).

\textsuperscript{18}The off-net-cost pricing principle dates back to Laffont et al. (2003), who found this pricing rule in a framework for Internet backbone competition.

\textsuperscript{19}The opportunity cost of stealing a caller away from the rival network is $c_0 + a = c + m$, whereas the opportunity cost of stealing a receiver away from the rival network is $c_0 - a = -m$. See López (2010) for a complete characterization of the off-net-cost pricing equilibrium.
firms set marginal prices at the opportunity cost of stealing the customers away from the rivals, and so marginal prices do not depend on market shares. The way consumers form expectations is then irrelevant for the level of the equilibrium marginal prices.

By setting calling and reception charges at the off-net cost, we have that \( \alpha_i = \frac{1}{2} - \sigma (F_i - F_j) \) (for \( i \neq j = 1, 2 \)). At equilibrium, market shares do not depend on expectations because there is full participation and, as commented above, usage prices are independent of market shares and symmetric. Thus, \( i \)'s profit can be rewritten as follows:

\[
\pi_i = \left( \frac{1}{2} - \sigma(F_i - F_j) \right) (F_i - f).
\] (10)

Solving the first-order conditions, we obtain the equilibrium fixed fees \( F_i = f + \frac{1}{2\sigma} \). The equilibrium profit is therefore \( \pi_i = \frac{1}{4\sigma} \), which is the profit that each network would obtain under unit demands. We also have that at equilibrium, profits are independent of the level of the access charge. As López (2010) points out, the reason is that all call activities yield zero profit: on-net calls cost (per unit) \( c \) and yield revenue (per unit) \( p + r = c \), originating an off-net call costs \( c_O + a \) while it yields revenue \( p = c + m = c_O + a \), and the cost of terminating an off-net call is \( c_O \) while it yields revenue \( a + r = a - m = c_O \).

If reception charges are restricted to be non-negative, the above analysis is only correct for \( m \leq 0 \), that is, for termination charges below the cost of termination. Suppose \( m > 0 \) and reception charges cannot be negative. Then it will be optimal to set reception charges at the minimum, i.e., \( r_1 = r_2 = r = 0 \). Hence, if termination charges are above cost, firms will not charge consumers for the reception of calls, even if they are allowed to do so. And the optimal call price will then be \( p_1 = p_2 = p = c + \frac{m}{2\sigma} \). In this case, call charges are again set at average marginal cost, but reception is "charged" (at zero) above the true cost of termination \( -m < 0 \). Firms now do make profits from traffic, in particular from terminating calls. Given the symmetry in call and reception charges, market share is again given by \( \alpha_i = \frac{1}{2} - \sigma(F_i - F_j) \). Hence, firms choose the fixed fee so as to maximize

\[
\pi_i = \alpha_i(mp)(p) + F_i - f
\] (11)

The first-order condition for a symmetric equilibrium now reads

\[
0 = -\sigma(mp)(p) + F - f + \frac{1}{2}
\]

so that \( F = f + \frac{1}{2\sigma} - mp(p) \) and equilibrium profit equals, again, \( 1/(4\sigma) \).\(^{20}\) We have the

\(^{20}\)This equilibrium exists and is unique when \( \sigma \) or \( m \) are not too high (see proof of Proposition 7 in Laffont et al. [1998a, Appendix B]).
following.

**Proposition 1 (Equilibrium)** (i) if $-\left(\frac{\beta}{1+\beta}\right) c < m < 0$, then as the noise vanishes there exists a unique symmetric equilibrium where marginal prices are set at the off-net cost $(p = c + m$ and $r = -m)$, $F = f + \frac{1}{2\sigma}$ and profit is neutral to the termination charge: $\pi = \frac{1}{4\sigma}$; (ii) if $m \geq 0$ and reception charges cannot be negative, then (for $\sigma$ and $m$ not too high) there exists a unique equilibrium in which $p = c + \frac{m}{2}$, $r = 0$, $F = f + \frac{1}{2\sigma} - mq(c + \frac{m}{2})$ and profit is neutral to the termination charge: $\pi = \frac{1}{4\sigma}$.

The profit neutrality result is independent of the exact specification of the randomness in the marginal utility for receivers. In particular, it holds even if noise does not vanish. However, in order to determine the socially optimal prices we will assume that noise vanishes in the following regular way (similar to the definition in JLT):

**Definition 2** A sequence of distributions $F_n(\varepsilon)$ with zero mean on domain $[\varepsilon, \tilde{\varepsilon}]$ is called regular if for any continuous function $h(\cdot)$ we have

$$\lim_{n \to \infty} E[h(\varepsilon)|\varepsilon \geq \varepsilon_0] = h(\varepsilon_0) \text{ for all } \varepsilon_0 \geq 0$$

and

$$\lim_{n \to \infty} E[h(\varepsilon)|\varepsilon \leq \varepsilon_0] = h(\varepsilon_0) \text{ for all } \varepsilon_0 \leq 0.$$  

It is straightforward to show that the optimal call and reception prices converge to $(p^*, r^*) = (\frac{c}{1+\beta}, \frac{\beta c}{1+\beta})$ when noise vanishes in a regular way. The intuition is that efficiency requires, in the limit, that the volume of calls $q$ satisfies $u'(q) + \bar{u}'(q) - c = 0$. Since sometimes callers will determine volume and sometimes receivers will determine the volume, the optimal prices are $(p^*, r^*)$. Notice that optimality requires a positive reception charge and therefore $m$ must be strictly negative. In fact, the socially optimal reception charge will be

$$m^* = -\frac{\beta c}{1+\beta}.$$

### 4 Termination-based price discrimination

In this section we allow the firms to set a fixed fee and (non-negative) prices for making and receiving calls that can depend on the network receiving and originating the call. That is, firm $i$ chooses $(F_i, p_i, r_i, \hat{p}_i, \hat{r}_i)$. We use the same set-up as in JLT (2004), except for the fact that we do not allow for negative reception charges and that we assume that consumers form expectations in a passive way.
Since we assume that all consumers subscribe to one of the networks, we can use the method of maximizing profits with respect to usage prices for making and receiving calls, keeping market share constant, by adapting the fixed fee accordingly. It is not surprising that the usage prices in a symmetric equilibrium candidate we find are the same as the ones found by JLT (2004) under the assumption that consumer expectations vary with respect to prices. For completeness, we include the analysis. We will again assume that there is randomness in the marginal utility of receivers and that this noise vanishes in a regular way.

We start the analysis with the market for on-net calls. It is optimal for network $i$ to maximize the size of the pie for on-net calls. The first-order conditions with respect to $p_i$ reads

$$\frac{d[U(p_i, r_i) + \tilde{U}(p_i, r_i) - cD(p_i, r_i)]}{dp_i} = 0,$$

while the one with respect to $r_i$ reads

$$\frac{d[U(p_i, r_i) + \tilde{U}(p_i, r_i) - cD(p_i, r_i)]}{dr_i} = 0.$$

As the noise vanishes, these equations can be solved to yield $p_i = c/(1 + \beta) \equiv p^*$ and $r_i = c\beta/(\beta + 1) = \beta p^* \equiv r^*$. (This is the same exercise as determining the socially optimal call and reception prices, as we did in the previous section.)

It is clear that there always exists an equilibrium with both off-net call and reception charges equal to infinity, so that no off-net calls will be made. This is independent of the level of the termination mark-up. Both networks then just offer efficient levels of on-net traffic and compete for subscribers by means of the fixed fees. When consumers expect both networks to be of equal size, then the equilibrium fixed fees will be equal to $f + 1/(2\sigma)$. Profits will be equal to $1/(4\sigma)$ for each firm. This type of equilibrium is pretty bad in generating consumer surplus as only on-net calls will be made. Clearly, if there are more than two networks this type of equilibrium is even worse.

We now solve for the optimal off-net call and reception charges in an equilibrium without connectivity breakdown. When consumers expect market shares to be equal, to keep true market shares constant at one half, network $i$ should adjust fixed fee as follows:

$$F_i = F_j + \frac{1}{2} \left\{ U(\hat{p}_i, \hat{r}_j) + \tilde{U}(\hat{p}_j, \hat{r}_i) - \hat{p}_i D(\hat{p}_i, \hat{r}_j) - \hat{r}_i D(\hat{p}_j, \hat{r}_i) - U(\hat{p}_j, \hat{r}_i) - \tilde{U}(\hat{p}_i, \hat{r}_j) + \hat{p}_j D(\hat{p}_j, \hat{r}_i) + \hat{r}_j D(\hat{p}_i, \hat{r}_j) \right\}. $$
The first-order derivative of profit with respect to \( \hat{p}_i \) reads

\[
\frac{1}{4} \left[ \frac{\partial U(\hat{p}_i, \hat{r}_j)}{\partial \hat{p}_i} - (c + m - \hat{r}_j) \frac{\partial D(\hat{p}_i, \hat{r}_j)}{\partial \hat{p}_i} - \frac{\partial U(\hat{p}_i, \hat{r}_j)}{\partial \hat{p}_i} \right].
\]

This can be rewritten as

\[
\frac{1}{4} \left[ \hat{p}_i - (c + m - \hat{r}_j) - \beta \hat{p}_i - E[\varepsilon \mid \varepsilon \geq \hat{r}_j - \beta \hat{p}_i] \right] \frac{\partial D(\hat{p}_i, \hat{r}_j)}{\partial \hat{p}_i}.
\]

The first-order derivative of \( \pi \) with respect to \( \hat{r}_i \), keeping market share constant at one half, reads

\[
\frac{1}{4} \left[ \frac{\partial U(\hat{p}_j, \hat{r}_i)}{\partial \hat{r}_i} + (\hat{p}_j + m) \frac{\partial D(\hat{p}_j, \hat{r}_i)}{\partial \hat{r}_i} - \frac{\partial U(\hat{p}_j, \hat{r}_i)}{\partial \hat{r}_i} \right].
\]

This can be rewritten as

\[
\frac{1}{4} \int_{-\beta \hat{p}_j}^{\hat{r}_i - \beta \hat{p}_j} \frac{1}{\beta} \hat{r}_i + \hat{p}_j + m - \frac{\hat{r}_i - \varepsilon}{\beta} q\left( \frac{\hat{r}_i - \varepsilon}{\beta} \right) f(\varepsilon) d\varepsilon,
\]

which in turn is equal to

\[
\frac{1}{4\beta} F(\hat{r}_i - \beta \hat{p}_j) E((\hat{r}_i + \hat{p}_j + m - \frac{\hat{r}_i - \varepsilon}{\beta}) q\left( \frac{\hat{r}_i - \varepsilon}{\beta} \right) \mid \varepsilon \leq \hat{r}_i - \beta \hat{p}_j).
\]

Let \( F^{(n)} \) represent a series of noise distributions that is regular according our definition. Let \( (\hat{p}^{(n)}, \hat{r}^{(n)}) \) denote the corresponding symmetric equilibrium candidate usage prices. By taking a suitable subsequence one may assume that either \( \hat{r}^n - \beta \hat{p}^n \leq 0 \) for all \( n \) or that \( \hat{r}^n - \beta \hat{p}^n \geq 0 \) for all \( n \).

Consider the first case. Then in the limit, as noise vanishes, the limit point \((\hat{p}, \hat{r})\) must satisfy \( \hat{r} - \beta \hat{p} \leq 0 \) and

\[
0 = (1 - \beta) \hat{p} - c - m + \hat{r} \quad \text{(12)}
\]
\[
0 = \hat{r} + m \quad \text{(13)}
\]

so that \( \hat{r} = -m \) and \( \hat{p} = (c + 2m)/(1 - \beta) \). The condition \( \hat{r} - \beta \hat{p} \leq 0 \) is satisfied if and only if \( m \geq -\beta c/(1 + \beta) \). (Note that these symmetric candidate equilibrium usage prices were reported in JLT (2004).)

Consider the second case next. Then in the limit, as noise vanishes, the limit point \((\hat{p}, \hat{r})\)
must satisfy $\hat{r} - \beta\hat{p} \geq 0$ and

$$
\begin{align*}
0 &= \hat{p} - c - m \\
0 &= \hat{r}(1 - 1/\beta) + \hat{p} + m
\end{align*}
$$

so that $\hat{p} = c + m$ and $\hat{r} = \beta(c + 2m)/(1 - \beta)$. The condition $\hat{r} - \beta\hat{p} \geq 0$ is satisfied if and only if $m \geq -\beta c/(1 + \beta)$. JLT (2004) discard this candidate equilibrium on the ground of second-order considerations.

In fact, the second-order derivative of profits with respect to $\hat{r}$ is strictly positive at the equilibrium candidate, also in the first case. Namely, at $\hat{r} = -m$

$$
\frac{\partial^2 \pi}{\partial \hat{r}^2_i} = \frac{1}{4}(1 - \frac{1}{\beta})q'(\hat{p})F(\hat{r} - \beta\hat{p})/\beta > 0.
$$

If $-m \leq 0$ then the optimal reception charge must be $\infty$, since increasing the reception charge will then improve profits. Hence, for $m \geq 0$ the only symmetric equilibrium is the one with connectivity breakdown. For $m < 0$ and thus $-m > 0$, the unique equilibrium candidate with $\hat{r} < \beta\hat{p}$ is the one with minimal reception charge, i.e. with $\hat{r} = 0$, and thus $\hat{p} = (c + m)/(1 - \beta)$.

Let us consider the fixed fee in a symmetric equilibrium candidate without connectivity breakdown. Because of passive expectations and symmetric usage prices, $\alpha_1 = \frac{1}{2} + \sigma(F_2 - F_1)$. Profit made from traffic on network $i$ equals $\alpha_i(1 - \alpha_i)\hat{R}$ where $\hat{R} = (\hat{p} - c)q(\hat{p})$. Network $i$ thus maximizes

$$
\alpha_i((1 - \alpha_i)\hat{R} + F_i - f).
$$

In a symmetric equilibrium we thus have $F_1 = F_2 = f + \frac{1}{2\sigma}$ and profits are maximized when $\hat{R}$ is maximized.

From the point of the firms, the profit maximizing termination mark-up thus satisfies $(c + m)/(1 - \beta) = p^M$ or $m = p^M(1 - \beta) - c$, if the latter expression exceeds $-c_T$ and is negative. For this to be feasible the call externality (that is, $\beta$) should be sufficiently strong. Bill and Keep $(m = -c_T)$ is optimal if $p^M(1 - \beta) - c < -c_T$. Otherwise the optimal $m$ will be the largest $m < 0$ for which an equilibrium without connectivity breakdown exists. We will denote this by $\hat{m}$.

From a social point of view, the optimal termination charge would be the one that makes the off-net call price equal to $p^* = c/(1 + \beta)$. This requires $m = -2\beta c/(1 + \beta)$. Clearly, this requires the regulator to know precisely the value of $\beta$. If the regulator faces uncertainty about this value, it may be a second-best but secure option to have Bill and Keep.
Proposition 3 (i) As the noise vanishes, the unique symmetric equilibrium candidate without connectivity breakdown has

\[ p = \frac{c}{1 + \beta}, \quad r = \frac{\beta c}{1 + \beta}, \quad \hat{p} = \frac{(c + m)/(1 - \beta)}, \quad \hat{r} = 0, \quad F = f + \frac{1}{2\sigma}, \]

yielding equilibrium profit

\[ \pi^* = \frac{1}{2\sigma} + \frac{1}{4}(\hat{p} - c)q(\hat{p}). \]

(ii) The profit maximizing termination mark-up satisfies

\[ m^* = \min\{\bar{m}, \max\{-c_T, p^M(1 - \beta) - c\}\}. \]

(iii) The social welfare maximizing termination mark-up equals

\[ m^{**} = \max\{-c_T, -2\beta c/(1 + \beta)\}. \]

5 Concluding remarks

We have analyzed how termination charges affect retail price competition when firms can charge consumers for receiving calls. Compared to earlier literature on this topic we assume that consumers form expectations about network sizes in a passive, but ex-post rational way. Moreover, we restrict reception charges to be non-negative. When firms cannot set different prices for on-net and off-net traffic, expectations over network sizes do not matter and we obtain the standard profit neutrality result in this case. Firms will set a positive reception charge only if termination charge is below termination cost. In this sense we confirm European operators’ warnings that further reductions in termination charges may end the Calling Party Pays Regime. This is not necessarily a bad thing in terms of social welfare. In fact, when receivers’ utility is random and thus receivers sometimes determine the call volume, it is optimal to have strictly positive reception charges. This can only be achieved by setting termination charge below cost. On the other hand, a symmetric equilibrium with positive reception charges only exists if receiver’s utility is sufficiently high. If receiver and caller derive the same benefit, and if termination cost constitutes half of the total cost of a call, then Bill and Keep leads to the socially optimal outcome. (DeGraba (2003) makes this point without formal model.) In this case call and reception charges are the same. This may resemble the situation in the US market pretty well.

When firms are allowed to distinguish between on-net and off-net traffic, we already know from Hurkens and López (2010) that the way expectations are formed are very important
under a CPP regime. Firms typically will prefer high termination charges while below or at cost termination charges are optimal from a social point of view. This is also true for relatively low levels of receiver utility. Only for very strong levels of call externalities firms would prefer below cost termination charges. The fact that most operators in Europe strongly oppose cuts in termination rates suggests that call externalities are not believed to be very strong.

Under the RPP regime we have shown that with termination based-price discrimination firms will charge on-net reception (since this maximizes the surplus from on-net traffic when there is vanishing noise in the receiver’s utility. However, extremely high off-net reception charges will often lead to connectivity breakdown. In particular, this must occur when the termination mark-up is nonnegative. For a negative termination mark-up, a symmetric equilibrium without connectivity breakdown may exist. It is characterized by zero reception charges off-net.

Our findings shed some light on how termination rates may affect the business model used in different countries. While in Europe high termination rates exist, it is optimal for the operators to stick to a CPP regime. Otherwise either connectivity breakdown becomes an issue (which hurts all consumers) or firms will be forced to agree not to use termination-based price discrimination. In both cases profits will be much lower. In countries as the US with low termination rates, it is optimal to charge consumers for receiving calls to as to recover some of the costs related to termination service. Firms may prefer to use termination-based price discrimination in this case, as it opens the possibility of an equilibrium with higher profits. However, it also creates the problem that connectivity breakdown may occur. And once firms stick to the RPP regime and do not use termination-based price discrimination, they have no incentive to try to manipulate the termination rates because of the profit neutrality results.

Apart from the difference between Europe and the US in termination rates and the regimes used, there is the matter of penetration, which is much higher in Europe. In order to address how penetration rates are related to termination rates and pay regimes, we need to allow for elastic subscription demand. We plan to do so in the near future.

**APPENDIX**

We introduce some notation and derive some useful derivatives.

Define

\[ D_{ij} = D(p_i, r_j) = [1 - F(r_j - \beta p_i)]q(p_i) + \int_{\varepsilon}^{r_j - \beta p_i} q\left(\frac{r_j - \varepsilon}{\beta}\right) f(\varepsilon) d\varepsilon, \]
\[ U_{ij} = U(p_i, r_j) = [1 - F(r_j - \beta p_i)]u(q(p_i)) \\
+ \int_{\xi}^{r_j - \beta p_i} u(q(\frac{r_j - \varepsilon}{\beta}))f(\varepsilon)d\varepsilon, \]

and

\[ \tilde{U}_{ij} = \tilde{U}(p_i, r_j) = \int_{\xi}^{\varepsilon} [\beta u(q(p_i)) + \varepsilon q(p_i)]f(\varepsilon)d\varepsilon \\
+ \int_{\xi}^{r_j - \beta p_i} [\beta u(q(\frac{r_j - \varepsilon}{\beta})) + \varepsilon q(\frac{r_j - \varepsilon}{\beta})]f(\varepsilon)d\varepsilon. \]

Then,

\[ \frac{\partial D_{ij}}{\partial p_i} = [1 - F(r_j - \beta p_i)]q'(p_i) \]

and

\[ \frac{\partial D_{ij}}{\partial r_j} = \int_{\xi}^{r_j - \beta p_i} \frac{1}{\beta} q'(\frac{r_j - \varepsilon}{\beta})f(\varepsilon)d\varepsilon. \]

Further,

\[ \frac{\partial U_{ij}}{\partial p_i} = [1 - F(r_j - \beta p_i)]p_i q'(p_i) = p_i \frac{\partial D_{ij}}{\partial p_i}. \]

and

\[ \frac{\partial U_{ij}}{\partial r_j} = \int_{\xi}^{r_j - \beta p_i} \frac{1}{\beta} \frac{r_j - \varepsilon}{\beta} q'(\frac{r_j - \varepsilon}{\beta})f(\varepsilon)d\varepsilon. \]

Finally,

\[ \frac{\partial \tilde{U}_{ij}}{\partial p_i} = \int_{r_j - \beta p_i}^{\varepsilon} (\beta p_i + \varepsilon)q'(p_i)f(\varepsilon)d\varepsilon \]

and

\[ \frac{\partial \tilde{U}_{ij}}{\partial r_j} = \int_{\xi}^{r_j - \beta p_i} r_j q'(\frac{r_j - \varepsilon}{\beta}) \frac{1}{\beta} f(\varepsilon)d\varepsilon = r_j \frac{\partial D_{ij}}{\partial r_j}. \]

References


