Efficiency of NGN Interconnection Charging Methods

Aleksandra Kostić-Ljubisavljević[1], Vesna Radonjić[2], Vladanka Aćimović-Raspopović[3] and Snežana Mladenović[4]

Abstract – **Telecommunications networks are being upgraded from current generation circuit switched technology to Next Generation all-IP networks. These new networks will have lower operating costs and offer opportunities for new services. This paper reviews whether current interconnect charging principles or Bill and Keep will be more likely to promote dynamic and static efficiency gains when applied to voice and messaging services on NGNs.**

Keywords – **NGN interconnection, Efficiency, interconnection charging**

I. Introduction

Traditional electronic communications networks used for voice and related services employ circuit- switched technology and guarantee end-to-end quality of service. By contrast, the Internet employs packet-switched technology which, although less expensive to operate, does so on a "best efforts" basis with no guarantee of quality. To capture the efficiency benefits of the Internet, but also offer the quality benefits of traditional networks, the communications industry is developing Next Generation Networks (NGNs) capable of carrying voice and data to acceptable levels of quality depending on the consumer service.

While the transition to NGNs is at its early stages, regulators have already begun to assess its consequences for regulating interconnection charges between network operators for traffic which passes between their networks. A focus of the regulatory debates has been whether the bill-and-keep (BAK) interconnection charging model (whereby no interconnection payments are made for either termination or origination) is preferable to other charging models, and in particular to the initiating party's network pays model (IPNP, under which the originating network pays a termination fee to the terminating network). BAK is seen as having a key advantage over IPNP: by eliminating termination charges, because it saves regulators from the resource-intensive and often contentious task of setting termination charges. IPNP, on the other hand, is viewed by many regulators as leading to excessive termination charges requiring price caps.

This paper is organized as follows. After short introduction in section I, the literature review on NGN interconnection is presented in section II. Section II is dealing with "hot potato" problem of interconnection that is relating to termination of call/message/connection in NGN networks. In section IV are factors which determine the efficiency of an interconnection model presented and analyzed. Section V is related to welfare consequences of inefficient interconnection. At the end of paper conclusion is placed.

[1]Aleksandra Kostić-Ljubisavljević is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia E-mail: a.kostic@sf.bg.ac.rs
[2]Vesna Radonjić is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia E-mail: v.radonjic@sf.bg.ac.rs
[3]Vladanka Aćimović-Raspopović is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia E-mail: v.acimovic@sf.bg.ac.rs
[4]Snežana Mladenović is with the Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia E-mail: snezanam@sf.bg.ac.rs

II. Literature Review on NGN Interconnection

There is little research on the charging principles most appropriate to NGN, though there has been some debate on the merits of BAK regardless of the underlying technology.

One of the first papers to study the economics of interconnection of all-IP networks is [1]. Author describes the Internet as having two distinct groups of subscribers, consumers and websites, who each gain value from more of the other side being present. Internet surfers gain more value if there are more websites of interest and websites gain value if there are more surfers. If a new surfer or website joins the network, there would be a positive benefit to those already connected to the network, but this benefit is not factored into the buying decision of the individual website or surfer. Therefore, in order to achieve the highest level of welfare across both sides of the market, one needs a pricing structure that encourages the highest number of users from both sides, which may be quite different for the two sides and need not necessarily relate to the costs caused by each party. The more flexible approach than simply charging at cost is required in all-IP network environments since it is extremely hard to calculate costs when multiple services are provided within a common network.

DeGraba in [2] proposed a BAK regime he termed Central Office Bill and Keep (COBAK). Central Office is the American equivalent of a local exchange. There was two rules proposed. First, the receiving party's carrier cannot charge an interconnecting carrier to terminate a call. Secondly, the calling party's carrier is responsible for the cost of transporting a call to the called party's central office. COBAK is proposed as a default rule for interconnection if carriers cannot agree on alternative terms in commercial negotiations.

The COBAK proposal is premised on three observations. First, that both parties generally benefit from a call, secondly that competition is more effective when carriers recover costs from their own customers and thirdly, that an arbitrage opportunity exists when regulation results in different charges being assessed for the same facility. The principle current benefit of COBAK is that it "significantly reduces" the terminating monopoly problem. Other benefits are that it will lead to more efficient pricing and therefore more efficient usage and that it reduces the need for regulatory intervention.

Responding to [2], Wright sets out two problems [3]. First, that COBAK fails to internalise network externalities between calling parties and secondly its failure to apply Ramsey principles. Arguments on the first objection is that the calling party receives a direct benefit as a result of the called party being willing to accept the call and that this benefit is likely to be larger than that flowing in the opposite direction. If the calling party pays for the costs of the receiving the call by called party, this will result in an efficient transfer between the two types of callers. By imposing BAK, this transfer will be eliminated.

DeGraba responds to [3] rejecting his criticisms of the COBAK proposal [4]. The first criticism is rejected on the basis that artificially increasing origination charges by having the caller pay for termination may cause inefficient substitution of low cost wireline call by higher cost wireless calls. The second criticism is rejected on the grounds that the COBAK proposal will still impose a higher cost on the calling party even if the benefits of the call are shared equally by both parties.

III. The "Hot Potato" Problem

An argument often used against BAK is the "hot potato" problem which arises because communications providers have an incentive to hand over traffic to another network for termination as close to the point of origin as possible, thereby reducing their own costs and maximising the costs of the terminating network. If the terminating network is not able to recover these costs then, it is claimed, the terminating network will under invest.

The European Regulatory Group (ERG) suggests that the problem could be overcome by requiring operators to have a reasonable minimum number of interconnection points for BAK to be applicable to that operator. As discussed above this would involve the regulator in determining the topology of points of interconnection. The ERG then points out that if operators had to increase their network size to be BAK partners for other networks, the investment involved could be inefficient if infrastructures are unnecessarily duplicated [5].

The hot potato problem can be removed by making the originating network responsible for the costs of transport all the way to the terminating network's central office, including the cost of transit if involved [2]. This proposal maintains the incentive for the originating network to build an efficiently sized network and also maintains the customer-supplier relationship between the originating and terminating network and thus the incentive for investment as costs can be recovered from the originating network.

The choice of interconnection charging models will be critical to achieving the efficiency gains available from shifting to IP technology in networks that enable the network operator to manage Quality of Service (QoS). In particular, interconnection charges are important factors that impact the efficiency of service provision and network investment incentives.

IV. Factors which Determine the Efficiency of an Interconnection Model

There are a number of key factors which determine whether a given interconnection model is efficient, and most relevant are: externalities, network costs, stability of market conditions, and traffic balance [6].

Externalities and network costs

Messages are jointly consumed and thus their costs are jointly caused by both the receiving as well as the initiating party. The economic role of interconnection fees arises out of this fundamental characteristic. By initiating a message the initiating party causes an externality for the receiving party ("message externality"). Similarly, by deciding to join a network, a consumer may create an externality (a "network externality" or "subscriber externality") for other subscribers.

These externalities can be privately compensated between the end parties (e.g. by taking turns to call each other). If they arise "on-net" between customers on a single network, then retail charges can internalize them, with the effect of increasing total demand. However, where traffic flow "off-net" (i.e. between networks) achieving the right structure of retail prices to internalize externalities relies on payments at the wholesale level between network operators, because these payments signal retail pricing incentives. Through this mechanism interconnection fees enable externalities between retail customers to be internalized - although it is useful to bear in mind that this occurs at an aggregate level rather than for every individual message or subscription decision. Efficient interconnection charges can be derived from two elements: the nature of the externalities and the distribution of costs between the interconnecting networks.

IPNP will also provide a greater disincentive than the other major interconnection models to nuisance messages being sent. This results from the fact that IPNP can be implemented so that the full cost of the message being sent is recovered from the party initiating the message whereas alternative models recover some or all of this cost from the party receiving the message. This suggests that while all of the major interconnection models can support efficient message exchange in some circumstances, there are clear circumstances (that may arise relatively frequently in practice) in which IPNP will be superior to the other models. The academic and regulatory interconnection discussion has given little weight to the implications of the duality between messages for which, on average, compensation for externalities between retail parties is possible and messages where this is not the case.

Efficient compensation for externalities is also the basis for the efficient structure of transit charges. Where transit interconnection is required, consumers do not have a direct relationship with the transit network operator and the efficient balance of charges between the retail parties must be achieved through transit payments between transit networks - which might take the form of cascading payments along the route taken by the message. Thus, even if transit is required, interconnection fees can be used to induce an efficient allocation of retail charges (subject to any practical restrictions on transit agreements such as those arising from the technical limitations of today's internet).

Traffic balance

If traffic between "peers" (i.e., networks which have the same cost structure and customer profile) is "balanced", then it might seem that the direction of payments is not relevant to efficiency, because any choice of interconnection fee would result in exactly the same (zero) net payment between networks. By avoiding the transactions costs of making offsetting payments, BAK could be efficient under these circumstances.

However, even if traffic between some peer networks appears balanced at a point in time, network operators typically have scope to influence this balance and their costs. Furthermore, the balance of traffic and/or costs is likely to be disturbed by evolving market conditions. Thus, whether BAK is indeed efficient in a situation of traffic balance depends on whether market factors could change the balance or peer status.

Stability of market conditions

Efficiency requires that the interconnection fee can be adjusted to respond to market changes. These changes can be due to exogenous factors, for example, the introduction of new services, which alter the typical distribution of benefits between the initiating and the receiving party. Change in market conditions can also be endogenous - that is, caused by the incentives established by the adoption of a specific interconnection model. In particular, network operators will have strategic incentives to avoid costs or to favourably alter the traffic balance where those actions do not change their interconnection payments.

It is clear that an efficient interconnection model requires flexibility to respond to both exogenous market changes as well as strategic actions by network operators. BAK, which by definition implies an interconnection fee which is always equal to zero, can only be efficient if the market conditions are such that the efficient interconnection fee is equal to zero (i.e. either when traffic is balanced or along the diagonal in Fig. 1) and if these conditions are stable.

V. WELFARE CONSEQUENCES OF INEFFICIENT INTERCONNECTION

Interconnection charges can cause inefficiencies due to a mismatch between an operator's incremental revenues and incremental costs associated with interconnection. One way that an operator can react to a mismatch is by adjusting its retail pricing model in order to increase its revenues through retail payments, which would reduce consumer welfare to the extent that the change in the retail prices departs from the efficient retail model. Alternatively (or in addition), an operator may react to inefficient interconnection charges through a number of cost-avoidance strategies.

Business bias results from a customer-group-specific mismatch between a network's incremental interconnection costs and incremental retail revenues. For example, a terminating network that is not able to cover at least its marginal costs from termination would have an incentive to target customers who initiate more traffic than they receive (e.g., outbound telemarketers) and avoid customers who mainly receive messages (e.g., inbound call centres).

Freely negotiating networks would deter inefficient business bias by linking interconnection fees to overall traffic profiles. However, customer targeting can be socially harmful. First, targeting tends to leave some customer demand unserved or underserved: each message requires origination and termination, and a disincentive to provide one of these services (by avoiding customers that have a relatively high propensity to either originate or terminate messages) would tend to suppress traffic below its socially optimal level. Second, as a result of the targeting, traffic might be carried by the networks that are better able to bias their business, rather than the most efficient networks.

Network structure bias (the "hot-potato problem") occurs when investments in particular network elements are not fully rewarded through interconnection fees and where operators have the ability to determine the point of interconnection. For example, if interconnection fees do not respond to the costs that networks incur, then an operator would have the incentive to reduce its own costs by locating points of interconnection close to its own customers. This could then lead to inefficient network design (e.g., underinvestment in the trunk network) or inefficient network operation (e.g., inefficient vertical separation between network elements, so that more traffic is carried as transit).

All the distortions discussed above result from the same principle - a mismatch between incremental costs and incremental revenues associated with specific interconnection services. As IPNP, BAK, and RPNP represent a continuum of interconnection fees, similar (at least in a qualitative sense) inefficiencies can arise under each model where they are applied in circumstances that are not consistent with efficient message exchange for that model.

BAK is efficient in two specific situations: where traffic is evenly balanced between peers and this balance cannot be disturbed by either strategy or change in market conditions or where traffic is not evenly balanced but the distribution of retail benefits exactly matches the distribution of costs between the originating and terminating networks. If BAK is applied outside these circumstances, it will lead to inefficient traffic and subscription decisions and/or to strategic cost-avoidance behaviour [7].

The direct effect of BAK is to change the way in which an operator's costs are recovered. In particular, BAK prevents the terminating network from receiving revenues from the initiating party's network and thus the only source of revenue to the terminating network is its own retail customers. The operator of the terminating network has several options to respond to the implementation of BAK. It could seek to recover its costs from the receiving party through a fee for receiving the message. While this option solves the cost recovery problem from the perspective of the network operator, it will lead to inefficient retail prices except in situations mentioned above. For example, if the receiving party usually rewards the initiating party for initiating messages from which it benefits (e.g. through monetary transfers, consumption decisions, or social interaction) but has less scope to punish parties who initiate nuisance messages, then imposing BAK will not improve the incentives to send socially beneficial messages. However, BAK will encourage the distribution of socially undesirable nuisance traffic as is illustrated by the experience in the USA where termination rates are set to zero or at relatively low rates and where customers are charged for receiving as well as sending text messages. As we discussed earlier, in this situation IPNP would achieve a more efficient traffic mix.

Another option to compensate for the absence of termination fees is to increase the price for bucket plans, flat rates ("all-you-can-eat" plans), or fixed access fees. The effects of such price adjustments on traffic depend on whether the price increase explicitly or implicitly affects the volume of messages initiated and received [8].

If the price increase affects a customer's incremental costs of receiving messages (e.g., through paying a higher fixed access charge in order to receive more messages) then traffic distortions are similar to those that would result from directly charging for receiving messages. If the price adjustment increases the customer's incremental costs of initiating messages (e.g., through tightening of bucket limits on initiating messages), this would distort price signals, because the parties who bear the termination cost of a particular message are involved in that message neither at the initiating nor the receiving end. As a consequence, this model would generally lead to inefficient traffic and subscription decisions.

In an NGN environment, BAK may also lead to operators terminating off-net traffic only at low-quality levels and reserving high-quality capacity for on-net traffic -that is, for traffic that generates revenues for QoS provision. Regulators might attempt to prevent this reaction by prescribing quality levels at which interconnection must occur. However, operators could try to mitigate such a requirement by reducing or delaying their efforts to prioritize traffic or to invest in other QoS capabilities in the first place. In a QoS-differentiated scenario, the inflexibility of interconnection fees under BAK would also aggravate network structure bias: operators would have an incentive to offer higher quality services while avoiding associated additional costs by locating interconnection points closer to their own customers [9].

IPNP will tend to be efficient in the following scenarios:

- all benefits of the message accrue to the initiating party;
- the share of the benefits accrued to the receiving party is small compared to share of network costs incurred by the terminating network; and/or
- benefits to the receiving party are mostly relevant in situations where individuals interact repeatedly or where a monetary transaction between the initiating and receiving party accompanies each message.

The scope of these scenarios is broader than the specific nature of circumstances in which a zero interconnection fee (i.e. BAK) is efficient. This is simply a result of the fact that IPNP encompasses a range of interconnection fees while BAK represents a single interconnection fee. Outside of the circumstances listed above and ignoring externalities (e.g. network externalities), IPNP is not efficient.

VI. Conclusion

There is no single interconnection model that is efficient across all circumstances. In order to enhance economic efficiency regulators should intervene at the network layer that is closest to the market failure. The charging principles applied to any new service should be decided upon taking into account which party is likely to derive most benefit from a call/message and may include IPNP, RPNP and BAK. Interconnection settlements depend on externalities, network costs, stability of market conditions, and traffic balance.

Acknowledgement

This work is partially supported by the Ministry of the Science and technological development of the Republic of Serbia with the project TR32025.

References

[1] K. Yoon, "Interconnection Economics of All-IP Networks", *Review of Network Economics*, vol. 5, no. 3, 2006.

[2] P. DeGraba, “Bill and Keep at the Central Office as the Efficient Interconnection Regime,” *OPP Working Paper Series* no. 33, FCC, 2000.

[3] J. Wright, “Bill and Keep as the Efficient Interconnection Regime?”, *Review of Network Economics*, vol. 1, pp. 54-60, 2002.

[4] P. DeGraba, “Bill and Keep as the Efficient Interconnection Regime?: A Reply”, *Review of Network Economics*, vol. 1, pp. 61-65, 2002.

[5] Project Team on IP-Interconnection and NGN, ERG (07)09, (2007), available at: http://www.cmt.es/es/publicaciones/anexos/ERG%2807%2909_rept_on_ip_interconn.pdf.

[6] M. Dodd, A. Jung, B. Mitchell, P. Paterson, and P. Reynolds, “Bill-and-Keep and the Economics of Interconnection in Next-Generation Networks”, *Telecommunications Policy*, vol 33, pp. 324–337, 2009.

[7] V. Radojičić, A, Kostić-Ljubisavljević, “Inteconnection of Next Generation Networks”, *Proceedings of XXVII PosTel*, pp 341-350, 2009.

[8] R. Cadman, *NGN Interconnection: Charging Principles and Economic Efficiency*, NGNuk, London, 2007.

[9] J. S. Marcus, *Interconnection on an IP-based NGN Environment, Global Symposium for Regulators*, Dubai, United Arab Emirates, 2007, available at: http://www.itu.int/ITU-D/treg/Events/Seminars/GSR/GSR07/Documents_presentations/Session_III%20Scott%20Marcus_interconnect.pdf.