Knieps, Günter; Stocker, Volker

Working Paper
Market driven network neutrality and the fallacy of a two-tiered Internet traffic regulation

Discussion Paper, Institut für Verkehrswissenschaft und Regionalpolitik, No. 149

Provided in Cooperation with:
Institute for Transport Economics and Regional Policy, University of Freiburg

Suggested Citation: Knieps, Günter; Stocker, Volker (2014) : Market driven network neutrality and the fallacy of a two-tiered Internet traffic regulation, Discussion Paper, Institut für Verkehrswissenschaft und Regionalpolitik, No. 149, Inst. für Verkehrswiss. und Regionalpolitik, Freiburg i. Br.

This Version is available at:
http://hdl.handle.net/10419/101318

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden. Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen. Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
Market Driven Network Neutrality and the Fallacy of a Two-Tiered Internet Traffic Regulation

by Günter Knieps* and Volker Stocker**

Discussion Paper
Institut für Verkehrswissenschaft und Regionalpolitik
No. 149 – August 2014

Abstract:
Within a Generalized DiffServ architecture entrepreneurial flexibility for building intelligent multipurpose traffic architectures enables the provision of a variety of tailored traffic services for a wide range of heterogeneous application services. In order to solve the entrepreneurial traffic capacity allocation problem, we propose an incentive compatible pricing and quality of service (QoS) differentiation model for the Generalized DiffServ architecture resulting in market driven network neutrality. Optimal allocation decisions based on the opportunity costs of capacity usage require that all relevant traffic classes are taken into account simultaneously, rather than 1) excluding traffic classes (by means of minimum traffic quality requirements), 2) prescribing a maximum or minimum number of traffic classes or 3) arbitrarily including parameter specifications for or levels of QoS which are not reflected by demand side. It is particularly important that the opportunity costs of capacity reservations for deterministic premium traffic classes are interrelated with subsequent non-deterministic traffic classes. As a consequence, every form of market split would be artificial.

* Günter Knieps, Professor of Economics, University of Freiburg, Institute for Transport Economics and Regional Policy, Platz der Alten Synagoge, 79085 Freiburg, Germany, email: guenter.knieps@vwl.uni-freiburg.de

** Volker Stocker, University of Freiburg, Institute for Transport Economics and Regional Policy, Platz der Alten Synagoge, 79085 Freiburg, Germany, email: Volker.Stocker@vwl.uni-freiburg.de

The electronic version of this paper is available under: https://portal.uni-freiburg.de/vw/files/files-publikationen/disk/disk149
1. Introduction

In Europe, the introduction of network neutrality regulation was considered a regulatory fallacy for a long time. The general conviction was that regulation of network-specific market power regarding access to local loops would be sufficient to guarantee competitive downstream Internet traffic services markets. A paradigm shift occurred in September 2013, when the European Commission issued a proposal including a net neutrality regulation (cf. EC 2013). The proposal is still going through the legislative procedure. On April 3rd 2014 it was approved (with some adjustments) by the European Parliament in the first reading. Especially articles 23 and 24 consider the implementation of a net neutrality regulation which, due to its nature as a regulation, would be applicable in all Member States. The goal is to establish a two-tiered market for traffic services, culminating in a regulatory split between a market for best effort Internet access services and a market for specialized services. Article 23(2) specifies the regulatory market split, allowing the provision of specialized services endowed with higher and guaranteed levels of traffic quality, as long as general best effort traffic quality of the public Internet is not impaired “in a recurring or continuous manner” (EC 2014, p. 51). Irrespective of the detailed specifications of the network neutrality regulations under debate, the downstream market for Internet traffic services shall fall under the competence of the regulators. It is to be expected that regulators will closely monitor Internet traffic service providers – they will make sure that there are sufficient traffic capacities allocated to public Internet traffic services in such a way that traditional best effort traffic quality is not seriously hampered by the provision of specialized services. Approval by the Council would pave the way for a rapid implementation of new rules regarding traffic management in the Internet.

While Europe has made a step towards a net neutrality regulation, the situation is different in the U.S. After some regulatory interventions by the Federal Communications Commission (FCC) in the past,¹ the D.C. Circuit has recently struck

¹ The most prominent cases are the Madison River Case in 2005 and the Comcast Case in 2008 (cf. Lee and Wu 2009).
down parts of FCC’s *Open Internet Order* after traffic service provider Verizon had challenged those. The Court decision allows traffic service providers to offer traffic services based on paid prioritization. In the meantime, the FCC has issued a new proposal for legislation on May 15th. Regarding specialized services, the FCC considers them to bear the potentials of being both beneficial to users by stimulating network investments and harmful as they might threaten the open nature of the best effort Internet. However, the issue of a regulatory market split is opened for debate (cf. FCC 2014).²

The aim of this paper is to take a closer look at the subject of the proper management of scarce infrastructure capacities by means of entrepreneurial price and quality differentiation, to provide a critical assessment of the proposed regulatory framework in Europe, and to derive some lessons for the U.S. We conclude that the introduction of the proposed regulation in Europe would seriously harm the future evolution of the Internet. A regulatory enforcement of best effort principles for broadband Internet as well as the regulatory control of the impact of specialized services on the (minimum) quality of Internet access services seriously interferes with consistent economically based active traffic management. The proposed regulation would not only conflict with the multipurpose approach of the *Internet Engineering Task Force* (IETF) integrating data transmission for real- and non-real-time application services, but would also result in a regulatory induced artificial market split. Such a market split would fundamentally disturb entrepreneurial search processes for solving the capacity allocation problem and hence the evolution of innovative quality differentiated Internet traffic service markets.

The question how unregulated traffic service providers solve the entrepreneurial traffic capacity allocation problem is analyzed by means of an adequate pricing and quality of service (QoS) differentiation model. It will be shown that the convergence towards all-IP-based traffic capacities capable of providing all possible required traffic qualities creates a single relevant market for traffic service

---

² For a further comparison of the network neutrality debate in Europe in the U.S., see Knieps and Stocker (2014).
provision. Central to the analysis, traffic quality is best described by the network performance parameters delay, jitter (variations in delay) and packet loss rate.\(^3\) Whereas by means of deterministic traffic quality the worst case of delay, jitter and packet loss can be controlled, stochastic traffic quality only gives relative guarantees by mean, statistical or probabilistic delay, jitter or packet loss (cf. e.g. Martin et al. 2004, p. 52). Optimal allocation decisions based on the opportunity costs of capacity usage require that all relevant traffic classes are taken into account simultaneously, rather than 1) excluding traffic classes (by means of minimum traffic quality requirements), 2) prescribing a maximum or minimum number of traffic classes or 3) arbitrarily including parameter specifications for or levels of QoS which are not reflected by demand side. It is particularly important that the opportunity costs of capacity reservations for deterministic premium traffic classes are interrelated with subsequent non-deterministic traffic classes. As a consequence, every form of market split would be artificial. Thus, in particular a regulatory market split fixing the status quo best effort quality of TCP/IP for the open Internet and thereby prohibiting QoS traffic classes for a substantial amount of data traffic would conflict with the endogenous entrepreneurial search for QoS.

The paper is structured as follows: In section 2 a brief overview of the increasing role of market driven QoS differentiation is provided. First, the role of the Internet protocol (IP) as the driver of convergence of telecommunication services (e.g. voice over IP [VoIP]) and broadcasting services (e.g. IPTV) based on all-IP transmission of data packets is demonstrated. There has been an evolution of different isolated networks specialized either for telecommunications or broadcasting services applying different logistics towards harmonized logistics in the context of all-IP multipurpose traffic architectures. As for all-IP-based service provision, heterogeneous traffic qualities become essential, TCP/IP’s inherent disability to reflect heterogeneous demand for traffic qualities is fundamentally challenged. Based on the evolutionary search for traffic infrastructures initiated by the IETF different entrepreneurial OoS differentiation strategies can be developed. In the following section 3 we give a rationale for the im-

---

\(^3\) We use the term traffic quality and quality of service (QoS) interchangeably.
plementation of price and quality differentiation in a Generalized DiffServ architecture by means of resource reservation and/or prioritization of data packets. Within a Generalized DiffServ architecture entrepreneurial flexibility for building intelligent multipurpose traffic architectures enables the provision of a multitude of tailored traffic services for a wide range of heterogeneous application services (cf. Knieps 2013). In particular, traffic services can be endowed with deterministic and stochastic quality guarantees with respect to traffic quality. A pricing model for the Generalized DiffServ architecture is presented in section 4. The pricing model in Knieps (2011a) only focusing on priority rules within a simple DiffServ architecture is extended in order to allow for the entrepreneurial choice of more general architectures taking into account bandwidth reservations within multipurpose traffic management. The basic principle of pricing based on the opportunity costs of traffic capacity usage ensures monotony in traffic quality and prices across different traffic classes. A new type of network externality is derived such that an increase in bandwidth reservation for deterministic traffic classes leads to a shift in relevant variable cost functions of subsequent lower non-deterministic traffic classes. Incentive compatibility is ensured and market driven network neutrality results, because the relevant interclass externalities based on capacity reservation or prioritization capture the relevant opportunity costs according to priority levels. In section 5, implications for a critical appraisal of current proposed regulation and legislation are derived from the model. Section 6 concludes.

2. The Increasing Role of Market Driven QoS

2.1. From Parallel Infrastructures towards All-IP

Within the last two decades, the emergence and evolution of the Internet has triggered and spurred a convergence process of the telecommunications, information technology and media sectors. In the course of this convergence process isolated single-purpose infrastructures have turned into multipurpose infrastructures capable of carrying both telecommunications and broadcasting services (cf. Knieps 2003, pp. 217ff.). We divide this convergence process into three stages.
During the first stage, communications and broadcasting services were traditionally provided over parallel single-purpose infrastructures. Circuit-switched voice telephony was provided over the plain old telephone system (POTS) infrastructure and cable, radio and satellite networks mainly provided broadcasting services. After the commercialization of the Internet in the 1990s, packet-switched narrowband Internet access services were provided alongside circuit-switched voice telephony over the POTS infrastructure. As prevalent analogue dial-up or digital ISDN-based Internet access technologies enabled rather low downstream and upstream data rates, relevant Internet application services (e.g. email or surfing the web) were rather homogenous with respect to required bandwidth and traffic quality.

In a second stage, technological progress resulted in broadband Internet access technologies initially complementing and later increasingly replacing narrowband Internet access. While xDSL access technologies and upgraded cable networks provided broadband access, combinations with optical fiber (e.g. VDSL or HFC) led to further increases in up- and downstream data rates. At the same time, mobile broadband access was enabled by 3G (cf. e.g. Valdar 2006). Alongside, innovation produced new application services like VoIP or IPTV. Driven by the widespread adoption of IP, IP-based substitutes for traditional voice telephony and broadcasting services could be provided irrespective of the underlying infrastructure, i.e. on a platform independent basis (cf. Frischmann 2001, p. 6). Formerly isolated single-purpose infrastructures had begun to converge towards IP-based multipurpose architectures capable of providing telecommunications and broadcasting services.5

During stage 3, convergence towards all-IP infrastructures has been reinforced by further advances in access technologies, mainly fiber and 4G. Resulting in-

---

4 ISDN = Integrated Services Digital Network; DSL = Digital Subscriber Line; VDSL = very high speed DSL; HFC = hybrid fiber coax; 3G = third generation mobile access technology.

5 For example, from their inception as Internet service providers, cable operators provided quality-guaranteed VoIP services as substitutes for traditional circuit-switched voice telephony services (cf. Hazlett and Wright 2011, pp. 27f.).
creases in data rates enable the simultaneous use of multiple IP-based application services (e.g. triple play VoIP, IPTV and web surfing). Platform independent provision of application services in an all-IP environment is ensured on the basis of heterogeneous multipurpose infrastructures. Instead of different networks specialized either for telecommunications, broadcasting or content delivery based on different logistics, convergence towards all-IP multipurpose traffic architectures leads to common logistics based on harmonized standards. A blueprint for all-IP networks and corresponding management of data traffic has been provided in the context of next generation networks (NGNs) (cf. e.g. ITU 2004). Based on a strict delineation between application services and transport, the main idea is that the integrated provision of different services over heterogeneous multipurpose infrastructures is efficient. A global trend towards all-IP infrastructures is observable. A fundamental challenge inherent to all-IP multipurpose infrastructures is to ensure the provision of IP-based voice and broadcasting services with (at least) equivalent quality as in formerly isolated single-purpose infrastructures. The corresponding service provision inevitably requires tailored and hence differentiated traffic services based on traffic management as heterogeneous traffic qualities become essential.

2.2. **The Challenges of Best Effort TCP**

Creating an environment of differentiated traffic services, however, is subject of controversy. Especially the debate about network neutrality in the Internet discloses fundamental disagreement about the adequate management and allocation of resources, i.e. Internet traffic capacities. Originating from an era of parallel isolated single-purpose infrastructures, this debate is basically about how data packets are or should be transmitted over the Internet. There is an ongoing controversy as to whether traffic service providers should be obliged by regulation to treat all data packets equal or whether they should be allowed to perform (certain kinds of) traffic differentiation (cf. e.g. Schwartz and Weiser 2008, p. 1). Treating all data packets equal is in compliance with the standard of TCP’s passive traffic management performed on a decentralized end-to-end basis by the communicating edges. Traffic service providers do accept this traffic manage-
ment and do not intervene with capacity allocation and average traffic quality results endogenously. Such TCP/IP-based best effort principles constitute the reference point for what can be understood as strict or technical network neutrality. In contrast, active traffic management endows traffic service providers with the competence to autonomously manage traffic and hence congestion within their networks. Deviating from TCP/IP-based best effort principles, capacity allocation and quality differentiation strategies are centrally implemented by traffic service providers. Basic tools for differentiation are prioritization and resource reservation strategies. Here, network neutrality in its strict sense is violated.

TCP/IP’s inherent disability to reflect heterogeneous demand for traffic qualities has been recognized for a long time and has resulted in best effort-compliant strategies to increase average traffic quality but also in the development of best effort compliant bypass strategies. In compliance with TCP/IP-based best effort principles, traffic service providers have some instruments to ensure high levels of average traffic quality. They can do so by imposing user restrictions (e.g. volume caps) or by excessive investments in traffic capacities, so called over-provisioning (cf. e.g. Wu 2003). While such measures focus on preventing degradation of traffic quality by avoiding congestion, tailored traffic qualities for sensitive application services as required in an all-IP context cannot be obtained. Overlay networks\(^6\) constitute best effort-compliant bypass strategies. They aim at mitigating insufficiencies inherent to the traditional TCP/IP-based best effort environment. While routing overlay networks (RONs) increase routing efficiency, content delivery networks (CDNs) enable pay-for-priority traffic services by caching content on strategically distributed nodes inside networks, thus reducing the distance data packets have to travel to end users. Corresponding business models represent best effort-compliant price and quality differentiation strategies. However, such bypass strategies can mitigate the insufficiencies of TCP/IP, but cannot ensure adequate provision of real-time VoIP or IPTV services.

\(^6\) Overlay networks are networks “on top” of the basic Internet providing additional functionality. For an overview of overlay networks, see Clark et al. (2006).
Required differentiated and ensured levels of traffic quality can only be provided by means of active traffic management. Heterogeneous traffic quality requirements can be taken into account by the provision of tailored traffic classes. Further, in case of congestion, TCP/IP-based best effort average traffic quality creates discrimination potentials. On the one hand, bandwidth-intense application services congest available traffic capacities while non-bandwidth-intense application services suffer from resulting poorer traffic quality. On the other hand, quality-sensitive application services are discriminated against by quality-tolerant application services (cf. Knieps 2011a, p. 27ff.). Here, active traffic management can provide tailored solutions. Moreover, the widespread provision of specialized services based on the same capacities as Internet traffic services emphasizes the necessity for differentiated traffic services. Specialized services are bundled IP-based services (e.g. IPTV) consisting of an application service based on tailored and quality-ensured specialized traffic services provided by means of active traffic management. For the provision of quality-sensitive application services within converged all-IP Internet infrastructures active traffic management is fundamental. As best effort TCP/IP and best effort-compliant strategies can only provide insufficient solutions, a migration to a market driven QoS differentiation in the Internet based on unrestricted entrepreneurial search processes for adequate differentiation strategies is inevitable. Only then can an integrated optimization of traffic capacities take into account heterogeneous demand for traffic qualities. In the following section, we illustrate how unregulated traffic service providers can solve the capacity allocation problem making use of market driven QoS differentiation. It will be shown that the convergence to-

---

BEREC (2012, pp. 4f.) defines specialized services as follows: “Specialised services are electronic communications services that are provided and operated within closed electronic communications networks using the Internet Protocol. These networks rely on strict admission control and they are often optimised for specific applications based on extensive use of traffic management in order to ensure adequate service characteristics. When the performance of specialised services provided as vertically integrated services is compared with Internet access service offers, only the underlying electronic communications service component of the specialised services will be considered, and not the application layer. Specialised services may interwork with the electronic communication on the Internet through gateways executing the admission control function.”
wards all-IP based traffic capacities capable of providing all required traffic qualities creates a single relevant market for traffic service provision.

3. A Rationale for Price and QoS Differentiation in a Generalized DiffServ Architecture

The optimal allocation of traffic capacities must be based on the opportunity costs of capacity usage and is reflected in the providers’ entrepreneurial decisions on the number and specification of traffic classes. Heterogeneity in demand for traffic quality is crucial for the entrepreneurial decisions on traffic classes and desired levels of traffic quality. In an all-IP environment, application services vary significantly with respect to their requirements to traffic quality. While traditional Internet application services like email are rather robust regarding distortions in traffic quality and do not require high or stable levels of traffic quality, interactive real-time application services such as VoIP or video teleconferences are rather sensitive to traffic quality distortions – especially jitter is problematic. Other application services like video streaming are sensitive to packet loss while broadcast video services require low jitter and low packet loss. Application services with similar traffic quality requirements can be grouped into classes (cf. Ash et al. 2010, p. 5; ITU 2011, pp. 12ff.; Babiarz et al. 2006, pp. 12f.). Providing adequate traffic quality for a pre-defined number of traffic classes (aggregates of flows belonging to similar application services) instead of on a per flow basis considerably reduces the complexity of optimal QoS differentiation.

Active traffic management constitutes a toolkit supplying traffic service providers with means to solve the capacity allocation problem in the spirit of market driven net neutrality. It allows the provision and control of any traffic quality required by application services including deterministic (i.e. absolute bounds in the sense of worst-case guarantees for delay, jitter and packet loss are met) or stochastic guarantees (i.e. relative guarantees represented by mean, statistical or probabilistic delay, jitter or packet loss) of pre-specified levels of traffic quality. Especially real-time application services benefit from traffic services endowed
with deterministic guarantees for traffic quality. However, as such guarantees are based on the reservation of traffic capacities, corresponding traffic services are more resource-consuming than those giving stochastic (e.g., for video streaming) or no guarantees (e.g., for email) (cf., e.g., Martin et al. 2004, p. 54). Architectures enabling the implementation of QoS differentiation schemes have been issued by the IETF in a number of requests for comments (RFCs). In line with our focus on all-IP infrastructures, the importance of the Internet as an integrated multipurpose network based on resource sharing has been recognized from the beginning (e.g., Braden et al. 1994, pp. 5f.). Inspired by the quality requirements of traditional circuit-switched telephone services, efforts to provide quality-equivalent traffic services allowing for VoIP coexistent with best effort traffic were made early. In combination with the resource reservation protocol, the Integrated Services architecture (IntServ/RSVP)\textsuperscript{8} allows for flow-based traffic quality differentiation based on resource reservation and admission control. Fine-granularity QoS allows deterministic guarantees for traffic quality on a per flow basis but exhibits scalability problems (cf., e.g., Chen and Zhang 2004, pp. 368f.; Martin et al. 2004, p. 51). A more scalable but coarser alternative for QoS differentiation is the Differentiated Services architecture (DiffServ).\textsuperscript{9} Data packets are aggregated into traffic classes receiving class-specific treatment within the DiffServ domain. The outcome is a top-down priority scheme between traffic classes. Resulting traffic quality is monotonic decreasing with traffic classes. Implementing Multiprotocol Label Switching with traffic engineering capabilities (MPLS-TE), DS-TE (DiffServ aware MPLS-TE) allows for a combination of prioritization, resource reservation and admission control, furthermore increasing routing efficiency (e.g., Chen and Zhang 2004, pp. 370ff.; Evans and Filsfils 2007; Babiarz et al. 2006; Martin et al. 2004, p. 72).\textsuperscript{10}

\textsuperscript{8} The most fundamental RFCs in that context are RFCs 2210 (cf. Wroclawski 1997a), 2211 (cf. Wroclawski 1997b) and 2212 (cf. Shenker et al. 1997).

\textsuperscript{9} The most fundamental RFCs in that context are RFCs 2474 (cf. Nichols et al. 1998) and 2475 (cf. Blake et al. 1998).

\textsuperscript{10} For a concise and detailed overview of quality differentiation architectures see Evans and Filsfils 2007.
The Generalized DiffServ architecture\(^\text{11}\) takes the DiffServ architecture as “envelope architecture” but allows for combinations of active traffic management strategies. Based on components standardized by the IETF the entrepreneurial task for traffic service providers is the choice and implementation of an architectural design for active traffic management depending on the market demand for heterogeneous traffic services as required for different application services. Depending on the specification, deterministic and stochastic guarantees with respect to traffic quality can be given for traffic classes. A logistics mix – i.e. a variety of active traffic management strategies – enables the provision of a variety of traffic services providing differentiated levels of traffic quality. Hence, different degrees and characteristics of heterogeneity in demand for traffic quality are reflected by innovative QoS differentiations. The Generalized DiffServ architecture gives traffic service providers entrepreneurial flexibility for building intelligent multipurpose traffic architectures capable of providing a multitude of tailored traffic services for a wide range of heterogeneous application services (cf. Knieps 2013).

In competitive service markets, market driven network neutrality results from entrepreneurial search processes for optimal QoS differentiation schemes based on active traffic management. In order to prevent arbitrage and to ensure incentive compatibility, resulting top-down traffic management between traffic classes and monotony in traffic qualities must be supplemented by a corresponding pricing scheme based on congestion fees reflecting the opportunity costs of network usage (cf. Knieps 2013). The market driven principle of pricing based on opportunity costs constitutes a relevant reference point for an economically desirable net neutrality concept. Only a price and quality differentiation strategy based on the opportunity costs of traffic capacity usage can be stable. In the context of a broadband Internet, efficient price differentiation must ensure the required quality differentiation reflecting heterogeneous demand for traffic quality. Regulatory requirements with respect to traffic management cannot perform this task. Rather, the implementation of a market driven net neutrality requires an entrepreneurial design of price and quality differentiation comprising all IP-
based data traffic in such a way that each application service is priced according to the opportunity costs of used traffic capacities. Only then will providers of traffic services act neutrally (i.e. non-discriminatory) vis-à-vis application services with different capacity requirements – there are no incentives to discriminate against application services causing high opportunity costs. Rather, market driven net neutrality is violated if all data packets are transmitted with equal priority (cf. Knieps 2011a). For further network economic analysis, the relevant reference point can only be an endogenously resulting, market driven net neutrality.

Figure 1: Schematic Illustration of Price and QoS Differentiation for a Generalized DiffServ Architecture

A schematic illustration of the implementation process of an adequate price and quality differentiation strategy based on the Generalized DiffServ architecture is given in Figure 1. In the following section, we propose a pricing scheme based on the opportunity costs of traffic capacity usage for the Generalized DiffServ architecture.

4. A Pricing Model for the Generalized DiffServ Architecture

As mentioned above, from a network economic perspective, only a price and quality differentiation strategy based on the opportunity costs of traffic capacity can be stable. In an all-IP multipurpose network architecture in the sense of NGNs, data traffic from several services (mainly voice, video and data) is handled on an IP basis. As average traffic quality based on TCP/IP’s passive traffic
management cannot provide required differentiations reflecting heterogeneous demands for traffic quality, the transition to an Internet architecture based on active traffic management for all data traffic is inevitable. In this spirit, based on quality differentiated Internet traffic services, the Generalized DiffServ architecture enables the provision of a multitude of heterogeneous application services. In the course of corresponding quality differentiation strategies, guarantees for traffic quality can be given for different aggregates of data traffic grouped into traffic classes. By means of prioritization and resource reservation and admission control deterministic and stochastic guarantees with respect to traffic quality can be given.

Knieps (2011a) has presented an incentive compatible price and quality differentiation strategy for a DiffServ architecture. In a scenario in which differentiated levels of traffic quality are provided in different traffic classes based on pure prioritization strategies, a pricing scheme based on the opportunity costs of additional data packets in higher classes which are strongly determined by the delays imposed on the data packets in lower classes (interclass externalities) is developed. However, resource reservation strategies and hence deterministic traffic qualities relevant in the analysis of the impact of specialized service provision on the public Internet could not be illustrated. As the provision of “public” Internet traffic services and specialized services require the same resources (traffic capacities), capacity allocation between these service types produces a rivalry situation. Introducing a pricing model based on the Generalized DiffServ architecture, we can illustrate the effect of a marginal increase in bandwidth reservation for the provision of specialized services on “public” Internet traffic services.

The introduction of a pricing model for the Generalized DiffServ architecture aims at the simultaneous solution of short-run problems regarding optimal pricing (ideally based on the opportunity costs of network usage) and long-run investment in traffic capacities. We consider a traffic service network consisting of traffic capacities (i.e. bandwidth with dimension \( w \)). There is rivalry in consumption of these capacities between \( n \) different traffic classes as they are all provided over the same physical infrastructure (resource-sharing). Based on an
entrepreneurial decision, traffic classes are designed to match heterogeneous demands for traffic quality and are based on active traffic management. A subset of traffic classes \((i = 1, \ldots, m\) with \(m \leq n\)) is designed to cater to demand for highly quality-sensitive application services (e.g. jitter-sensitive real-time application services such as video conferencing). Similar to specialized services, they provide deterministic guarantees for pre-specified levels of (minimum) traffic quality. This is achieved by resource reservation and strict admission control. As degrees and characteristics of quality-sensitivity may vary between application services, traffic service providers might choose to provide several deterministic traffic classes differing in levels (e.g. lower vs. higher bounds for jitter) and characteristics (e.g. low jitter vs. low packet loss rate) of guaranteed traffic quality (cf. e.g. Ash et al. 2010). The higher the levels of deterministic traffic quality, the more resources must be used. We model the provision of such deterministic traffic classes by introducing virtual separation of capacities, i.e. a share of bandwidth is exclusively reserved for those traffic classes.

The bandwidth reserved for traffic class \(i\) is denoted by \(w_i\) with \(0 \leq w_i \leq w\) for all \(i = 1, \ldots, m\). Hence, the share of reserved bandwidth for deterministic traffic classes is \(0 \leq \sum_{i=1}^{m} w_i \leq w\). The residual capacity \(w - \sum_{i=1}^{m} w_i\) is available for traffic belonging to lower non-deterministic traffic classes. Capacity consumption of lower traffic classes is thus limited, as traffic capacities are partially dedicated to deterministic traffic classes. We consider the case where data packets from non-deterministic traffic classes \(j = m + 1, \ldots, n\) follow strict top-down priority scheduling. Hence, data packets belonging to higher traffic classes are prioritized vis-à-vis data packets from lower traffic classes. Traffic classes \(m + 1, \ldots, n - 1\) can be considered as providing stochastic guarantees for traffic quality based on statistic probabilities while traffic class \(n\) represents a best effort traffic class.\(^{12}\) Traffic classes are chosen according to a ranking taking into account traffic quality. The top priority class is the closest substitute to a deter-

\(^{12}\) The term best effort is chosen here because average traffic quality endogenously results in this class depending on available capacity and demand for traffic services. However, as the Generalized DiffServ architecture is based on active traffic management, this traffic class is different from an average traffic quality within a best effort TCP/IP-based Internet resulting from passive traffic management performed by the communicating edges.
ministic traffic class with a statistically guaranteed absence of queueing delay, jitter and packet loss. In the context of an architecture based on pure prioritization between traffic classes the opportunity costs of traffic capacity usage can be described by taking the concept of interclass externalities as a reference point. In order to include the role of specialized services we extend the model to fit the context of a multipurpose Generalized DiffServ architecture. We apply a more general perspective of rivalry for traffic network resources used for the provision of different traffic classes. Referring to the European regulation ante portas, we can illustrate the effect of a marginal increase in bandwidth reservation for the provision of specialized services on “public” Internet traffic services.

The marginal extension of reserved capacity $w_i$ and resulting externality effects on lower non-deterministic traffic classes indicate opportunity costs. While the allocation of dedicated capacities for traffic classes $i = 1, ..., m$ is assumed to be an ex ante entrepreneurial decision, capacity allocation between lower traffic classes $j = m + 1, ..., n$ results endogenously on residual capacity $w - \sum_{i=1}^{m} w_i$.

The inverse demand function for bandwidth reservation in traffic class $i$ does not vary over time and is denoted by $P_{it} \equiv P_i(w_i)$. Let $Q_{jt}$ denote the amount of data traffic in period $t$ in traffic class $j$ with $j = m + 1, ..., n$. The inverse demand function for packet transmission in traffic class $j$ is denoted by $P_{jt}(Q_{jt})$. As we do not aim to analyze intertemporal demand interdependencies, we assume demand to be independent across time periods. Moreover, we assume zero income effects and exclude re-shifting of traffic capacity between time periods. The costs of traffic capacity are denoted by $\rho(w)$. Modeling a multipurpose architecture, gains from multiplexing are ensured as all traffic services are provided over the same traffic capacities.\(^{13}\) Let $k_{jt}(Q_{m+1t}, ..., Q_{jt}, w - \sum_{i=1}^{m} w_i)$ be the private (average) variable costs of packet transmission in traffic class $j$ given residual capacity $w - \sum_{i=1}^{m} w_i$. Private (average) variable costs of data traffic vary with the degree of quality-sensitivity of corresponding application services. For non-deterministic traffic classes $j = m + 1, ..., n$ the following must hold:

\(^{13}\) E.g. in the case of two traffic classes with traffic class 1 being based on resource reservation $\rho(w) < \rho(w_1) + \rho(w - w_1)$.\)
• With constant capacity, an increase in traffic flows in traffic channel \( j \) in period \( t \) slows down any data packet belonging to the same traffic class, i.e.

\[
\frac{\partial k_j(t; w - \sum_{i=1}^{m} w_i)}{\partial Q_j(t)} > 0 \quad \forall \ j = m + 1, \ldots, n,
\]

and those belonging to down-stream traffic classes

\[
\frac{\partial k_k(t; w - \sum_{i=1}^{m} w_i)}{\partial Q_j(t)} > 0 \quad \forall \ k \neq j \ with \ k > j.
\]

• Thus, resulting marginal externality costs consist of intraclass externalities

\[
\frac{\partial k_j(t; w - \sum_{i=1}^{m} w_i)}{\partial Q_j(t)} \cdot Q_j(t) > 0
\]

and interclass externalities

\[
\sum_{k=j+1}^{n} \frac{\partial k_k(t; w - \sum_{i=1}^{m} w_i)}{\partial Q_j(t)} \cdot Q_k(t) > 0.
\]

Extending the interclass externality pricing model, we introduce the effects of capacity reservation for the provision of deterministic traffic classes \( i = 1, \ldots, m \). As in those traffic classes a certain minimum level of traffic quality is guaranteed on a deterministic basis, private (average) variable costs are usage-independent and hence negligible. A marginal increase in reserved capacity creates an interclass externality on all lower traffic classes with no or non-deterministic traffic quality guarantees:

\[
\frac{\partial k_j(t; w - \sum_{i=1}^{m} w_i)}{\partial w_i} > 0 \quad \forall i = 1, \ldots, m \ and \ j = m + 1, \ldots, n.
\]

Capacity reservation for deterministic traffic classes reduces residual capacity for each non-deterministic traffic class. Thus, a marginal increase in bandwidth reservation for traffic class \( i \) leads to an upward shift in the variable cost functions in traffic classes \( j = m + 1, \ldots, n \). As a consequence, interclass externalities from capacity reservation are strictly higher than interclass externalities from prioritization. A corresponding pricing scheme must take the opportunity costs of traffic capacity reservation, or traffic capacity consumption respectively, as reference point. We consider the case of competition where each traffic service provider chooses channel capacities and packet prices in each traffic class.
The welfare maximization is defined by the following optimization problem:

\[
\max S(Q_{m+1t}, \ldots, Q_{nt}, w_1, \ldots, w_m, w)
\]

\[
= \sum_{i=1}^{m} \int_0^{w_i} P_i (\hat{\omega}_i) d\hat{\omega}_i + \sum_{t=1}^{T} \sum_{j=m+1}^{n} \int_0^{Q_{jt}} P_jt (\hat{Q}_{jt}) d\hat{Q}_{jt}
\]

\[
- \sum_{t=1}^{T} \left[ k_{m+1t} \left( Q_{m+1t}, w - \sum_{i=1}^{m} w_i \right) Q_{m+1t} + \ldots \right]
\]

\[
+ k_{nt} \left( Q_{m+1t}, \ldots, Q_{nt}, w - \sum_{i=1}^{m} w_i \right) Q_{nt} \right] - \rho(w)
\]

Necessary conditions for the resulting welfare maximum can be derived by differentiating w.r.t. \(w_1, \ldots, w_m, Q_{m+1t}, \ldots, Q_{nt}\) for each \(t = 1, \ldots, T\) and w.r.t. \(w\). We can derive optimal pricing rules based on the opportunity costs of bandwidth reservation or packet transmission in each traffic class. Optimal congestion fees illustrate these opportunity costs.

- **Deterministic traffic classes \(i=1,\ldots,m\):**
  The optimal price for traffic class \(i\) is determined by negative externalities on lower non-deterministic traffic classes. Irrespective of actual traffic flows (i.e. usage levels) these interclass externalities solely depend on the share of reserved bandwidth \(w_i\). A marginal increase in bandwidth \(w_i\) results in upwards shifts in all variable cost functions in non-deterministic traffic classes \(j = m + 1, \ldots, n\).\(^{14}\)

\[
\tau_i = P_i = \sum_{t=1}^{T} \sum_{j=m+1}^{n} \frac{\partial k_{jt}(\cdot, w - \sum_{i=1}^{m} w_i)}{\partial w_i} \cdot Q_{jt}
\]

\[\text{with} \quad \frac{\partial k_{jt}(\cdot, w - \sum_{i=1}^{m} w_i)}{\partial w_i} > 0 \ \forall \ j = m + 1, \ldots, n ; \ t = 1, \ldots, T\]

\(^{14}\) Notice that there are no interclass externalities on lower deterministic traffic classes as traffic quality in those traffic classes is guaranteed on a deterministic basis and hence independent of resource reservation for other traffic classes. This is represented by \(0 \leq \sum_{i=1}^{m} w_i \leq w\).
• **Non-deterministic traffic classes j=m+1,..,n-1:**
The optimal congestion fee for non-deterministic traffic class j depends both on intraclass and interclass externalities.

\[
\tau_{jt} = P_{jt} - k_{jt} \left( \cdot, w - \sum_{i=1}^{m} w_i \right) \\
= \frac{\partial k_{jt}(\cdot, w - \sum_{i=1}^{m} w_i)}{\partial Q_{jt}} \cdot Q_{jt} + \sum_{k=j+1}^{n} \frac{\partial k_{kt}(\cdot, w - \sum_{i=1}^{m} w_i)}{\partial Q_{jt}} \cdot Q_{kt}
\]

Intraclass externalities \hspace{2cm} Interclass externalities

\[ t = 1, \ldots, T \]

• **Traffic class n:**
The optimal congestion fee for the lowest traffic class solely depends on intraclass externalities.

\[
\tau_{nt} = P_{nt} - k_{nt}(\cdot) = \frac{\partial k_{nt}(\cdot, w - \sum_{i=1}^{m} w_i)}{\partial Q_{nt}} Q_{nt}
\]

\[ t = 1, \ldots, T \]

• **Optimal investment rule:**

\[
\frac{\partial \rho}{\partial w} = - \sum_{t=1}^{T} \sum_{j=m+1}^{n} \left[ \frac{\partial k_{jt}(\cdot, w - \sum_{i=1}^{m} w_i)}{\partial w} \cdot Q_{jt} \right]
\]

Due to deterministic traffic qualities, there is no benefit from capacity expansion and thus investment in traffic classes \( i = 1, \ldots, m \). A marginal increase in reserved bandwidth \( w_i \) cannot generate benefits to traffic class \( i \) but would rather constitute a waste of resources.

Solving equations (1) - (4) simultaneously results in optimal allocation of traffic flows \( Q_{m+1t}^*, \ldots, Q_{nt}^* \ \forall \ t = 1, \ldots, T \). We can also determine optimal capacity dimension \( w \). Capacity extension is desirable until the marginal cost of capacity
extension is equal to the marginal benefits of reduced opportunity costs of traffic capacity usage in each traffic class.

While intraclass externalities indicate traffic quality within each traffic class, interclass externalities indicate opportunity costs caused by traffic capacity usage in one class with a negative effect on subsequent lower traffic classes. Hence, adequate price differentiation strategies for traffic quality differentiation based on pure prioritization reflect interclass externalities. In the case of a traffic class with dedicated traffic capacities and hence deterministically guaranteed traffic quality, however, actual capacity usage is not the adequate measure for resulting interclass externalities. In fact, capacity reservation translates into lower residual capacity for lower non-deterministic traffic classes and thus a new form of interclass externalities serve as adequate reference point for the opportunity costs of bandwidth reservation. Intraclass externalities in traffic class $i = 1, \ldots, m$ are irrelevant as pre-specified minimum levels of traffic quality are guaranteed irrespective of actual traffic flows (i.e. capacity usage) on a deterministic basis. Tailored levels of traffic quality provided in deterministic traffic classes are sufficient for relevant application services. Due to strict priority scheduling, data packet transmission in non-deterministic traffic class $j$ has negative effects on lower traffic classes in such a way that interclass externalities reflect corresponding opportunity costs. Depending on the demand structure, intraclass externalities are (probably rather) irrelevant as we expect traffic quality in traffic class $j$ to still be sufficiently high for relevant application services. Data packet transmission in traffic class $n$ solely produces intraclass externalities. Resulting congestion fees are monotonic, i.e. $\tau_{1t} > \cdots > \tau_{mt} > \tau_{m+1t} > \cdots > \tau_{nt}$. The resulting competitive pricing strategy is incentive compatible, i.e. consumers can self-select a traffic class according to their preferences while traffic services with higher traffic quality belonging to higher traffic classes are more expensive than traffic services with lower traffic quality provided in lower traffic classes. Both, traffic quality and prices are monotonically decreasing with traffic classes.

The effect of specialized services on public TCP/IP-based best effort Internet can be illustrated with an example. Assume two traffic classes. Deterministic
traffic class 1 is based on bandwidth reservation $w_1$ and gives deterministic guarantees for traffic quality. One could think of a specialized service for VoIP or IPTV. The provision of non-deterministic traffic class 2 is based on residual capacity $w - w_1$ and produces an average traffic quality. This scenario illustrates the market split proposed by the European Commission. Illustrating the impact of specialized (traffic) services on public best effort Internet traffic services, we consider the effect of a marginal increase in $w_1$ on the public Internet
\[
\frac{\partial k_{2t}(Q_{2t}, w-w_1)}{\partial w_1} > 0.
\]
The externality produced by the provision of specialized services is represented by an upward shift in the variable cost function $k_{2t}(Q_{2t}, w-w_1)$. It is apparent that rivalry for the same traffic capacities necessarily creates interdependencies and hence externalities between specialized services and Internet traffic services. Adequate QoS differentiation, however, should not result from regulatory traffic management, but market driven from entrepreneurial search processes based on active traffic management in the context of a Generalized DiffServ architecture.

5. The Fallacies of a Regulatory Market Split

It is well known from the disaggregated regulatory framework of network economics that the markets for network services are disciplined by active or potential competition. Due to the absence of irreversible costs potential competition is workable, even if advantages from bundling and subsequent economies of scale and scope exist. Competition on the markets for transport services in general is workable – in particular, this is true for traffic service markets for data packet transmission. Competition on downstream service markets is achieved by a disaggregated regulation of upstream monopolistic bottleneck components. Hence, network-specific market power is disciplined at its roots and cannot be leveraged into downstream service markets (cf. Knieps and Zenhaeusern 2008, pp. 127 ff.). Any form of market power regulation of traffic service markets is thus not only superfluous but detrimental. Nevertheless, as with any other service markets, general competition law and consumer protection laws should be applied.
The question whether regulation must protect providers of Internet application services from the abuse of market power by owners of upstream infrastructure (cf. Economides 2008, p. 210) can be answered in the negative. In these markets, active and potential competition can unfold between providers of Internet traffic services. The application of general competition law and consumer protection is then sufficient for ensuring workable competition on service markets (cf. e.g. Faratin et al. 2007; Knieps and Zenhauesern 2008). As competition on the downstream markets for Internet traffic services is workable, there is no justification for prohibiting providers of traffic services from certain price and quality differentiation strategies. Such regulation restricts entrepreneurial freedom and constitutes a false positive regulatory fallacy, i.e. over-regulation. Rather, rich innovation potentials in the field of Internet application services can only unfold and be fully exhausted if heterogeneous levels of traffic quality required by heterogeneous application services can be provided instead of being restricted by regulation of traffic management.

Any form of regulatory intervention into entrepreneurial traffic management constitutes over-regulation, disturbing potentials for market driven solutions. Regulatory market splits conflict with entrepreneurial freedom to implement a market driven QoS architecture and subsequent active traffic management. In the following, we consider two forms of regulatory market splits which are currently under debate: (1) minimal traffic quality regulation (2) separation of best effort TCP/IP from specialized traffic services.

(1) Regulation of minimal traffic service quality fundamentally conflicts with the entrepreneurial choice of traffic service classes and the implementation of incentive compatible pricing strategies. As a consequence, there is reason to fear that a demand for low traffic quality is confronted with excessively high regulatory enforced minimum traffic quality. Rather, socially desirable application services – e.g. aiming to solve universal service problems – may be quality-sensitive, requiring tailored traffic services with guaranteed traffic quality. Instead of enforcing high levels of minimum traffic quality by regulation, required premium traffic services based on active traffic management should evolve from
entrepreneurial search processes and could be subsidized (cf. Knieps 2011b, pp. 17ff.).

(2) In the course of the current debate one important issue is whether from a regulatory perspective specialized traffic services are to be considered “outside” the public Internet and hence exempt from “Internet rules” or best practices. However, the provision of specialized services cannot be considered isolated and thus outside the public Internet. Rather, they are necessarily provided inside the Internet, based on a common resource pool. Any IP-based data transmission ultimately requires the use of the same traffic capacities, irrespective of which application services they are serving as inputs for. An adequate pricing model reflects rivalry in consumption for scarce traffic capacities within the entire market for traffic services – a market split into best effort Internet traffic services and specialized services becomes meaningless.

Extending the model of interclass externality pricing, the opportunity costs of usage of traffic service capacities serve as a reference point for the entrepreneurial pricing decision. As could be shown in section 4, the optimal capacity dimension \(w\) is (given the demand for bandwidth reservation) solely determined by the marginal benefits of capacity expansion in non-deterministic traffic classes. Further, it was emphasized that the provision of specialized services inevitably produces externality costs on lower non-deterministic traffic classes. In our model, the opportunity costs of a marginal increase in bandwidth reservation \(w_i\) for deterministic traffic classes \(i\) on subsequent non-deterministic traffic classes \(j = m + 1, \ldots, n\) are reflected by upward shifts in private (average) variable cost functions \(k_{jt}(\cdot, w - \sum_{i=1}^{m} w_i)\).

---

15 BEREC notices: “Both service categories [specialized services and Internet access services] usually share the same physical infrastructure and, depending on the ISP’s decision, the capacity is divided between the two when they are configured.” (BEREC 2012, p. 7; parenthesis added by authors). The FCC states in its latest proposal: “In the Open Internet Order, the Commission recognized that broadband providers may offer “specialized services” over the same last-mile connections used to provide broadband service” (FCC 2014, p. 22).
6. Conclusions

As competition between Internet traffic (access) service providers is workable, the question raised in the course of the current net neutrality debate whether traffic service providers should be allowed to grant themselves or selected providers of application services prioritized transmission of data packets within their networks is irrelevant from a regulatory policy perspective. There are no incentives for discrimination. In view of revenues, traffic service providers are indifferent between the provision of high-quality traffic services based on which high-quality application services are offered by third-party providers and providing equivalent ‘bundled’ application services themselves.

While Internet traffic services should be generally unregulated – as with any other service markets general competition law and consumer protection laws should be applied – the Commission’s current proposal stipulates a net neutrality regulation restricting active traffic management via economically desirable price and quality differentiation strategies. From a regulatory policy perspective such regulation of traffic services not only contradicts the fundamental principle of liberalized service markets, it also constitutes an over-regulation significantly restricting entrepreneurial search processes for innovative price and quality differentiation strategies by the providers of Internet traffic services. Moreover, the regulatory market split may give traffic service providers incentives to take advantage of unclear distinctions between service categories. The task of the regulator should be exclusively restricted to upstream local telecommunications infrastructure as long as there are no alternative network infrastructures available.

From a network economic perspective, only a price and quality differentiation strategy based on the opportunity costs of traffic capacity usage can be stable. Fundamental to converged multipurpose network architectures is that the complete data traffic originated from several services (voice, video and data) is handled on an all-IP basis. As traditional TCP/IP-based best effort passive traffic management cannot provide required differentiations reflecting heterogeneous demands for traffic quality, the transition to an Internet architecture enabling market driven QoS differentiation based on active traffic management for all
data traffic is inevitable. In this spirit, based on quality differentiated Internet traffic services, the Generalized DiffServ architecture enables the provision of a variety of heterogeneous application services. In the course of corresponding quality differentiation strategies, guarantees for traffic quality can be given for different aggregates of data traffic which are grouped into traffic classes. By means of prioritization and resource reservation, specific levels of traffic quality can be guaranteed on a deterministic or stochastic basis. Taking this into account, a market driven quality and price differentiation model is developed.

The regulatory market split in TCP/IP-based best effort traffic services in the public Internet and quality-ensured specialized services is artificial and hampers entrepreneurial search processes for innovative architectures ensuring the efficient provision of tailored traffic services reflecting heterogeneous demand for traffic qualities.

References


BLAKE, S. ET AL. (1998), An Architecture for Differentiated Services, RFC 2475


WROCLAWSKI, J. (1997a), The Use of RSVP with IETF Integrated Services, RFC 2210.


Als Diskussionsbeiträge des
Instituts für Verkehrswissenschaft und Regionalpolitik
Albert-Ludwigs-Universität Freiburg i. Br.
sind zuletzt erschienen:


141. G. Knieps, P. Zenhäusern: The reform process of the railway sector in Europe: A disaggregated regulatory approach, December 2011


149. G. Knieps, V. Stocker: Market Driven Network Neutrality and the Fallacy of a Two-Tiered Internet Traffic Regulation, Paper to be presented at the 42rd Research Conference on Communication, Information and Internet Policy (TPRC), September 12-14, 2014, George Mason University, Arlington, VA