Raising Total Capacity Provision of Mobile Network Operators Through Throttling

Oliver Zierke*

First version: Oktober 18, 2018
This version: August 20, 2019

Abstract

In recent years mobile network operators worldwide introduced mobile plans that restrict data transmission rates of certain content types such as videos. However, this practice, known as throttling, is viewed critically as it violates the so-called net neutrality and is suspected to be used as an instrument to avoid network capacity improvements. Thus, this paper investigates the impact of throttling on mobile network operators' incentives to invest in the capacity of their network. A stylized game-theoretic model is used with a monopoly mobile network operator providing access to the internet for a homogeneous group of consumers. It is shown that throttling helps to use the capacity of mobile networks more efficiently, which, surprisingly, does not result in mobile operators offering less overall network capacity. Instead, the ability to throttle increases the incentives for them to improve their network capacity.

1 Introduction

In recent years, mobile network operators (MNO) in the US started to introduce new unlimited data plans without any data caps. Instead, companies introduced restrictions on the data transmission rate that could be applied to individual applications (Welch, 2018). For example, T-Mobile USA provides a tariff in which all video streams are restricted to the use of a maximum data transmission rate of 1.5 Mbps (T-Mobile USA, 2018). Other examples for MNOs providing tariffs with throttling the data transmission rates of certain content

*Chair of Internet and Telecommunications Business, University of Passau, Dr.-Hans-Kapfinger-Str. 12, 94032 Passau, Germany. Email: oliver.zierke@uni-passau.de.

types are Telkom LIT in South Africa (Saunders, 2018) or U Mobile in Malaysia (Qishin, 2018).

Deutsche Telekom's attempt to introduce a tariff with a similar restriction on data transmission rates in Germany failed due to resistance from the national regulator with a signal effect for the whole European Union (EU). The BNetzA found that throttling practices violate the net neutrality regulation of the EU (Meyer, 2017). This regulation states that "providers of Internet access services shall treat all traffic equally" (EU regulation 2015/2120). Scientific literature on net neutrality has shown that such regulation can be detrimental for overall welfare as it prohibits practices such as charging for prioritization of data streams in telecommunication networks, although such a paid prioritization often results in a better resource allocation (Krämer et al., 2013). The analysis of this paper shows that there are also reasons to doubt if a prohibition of throttling data streams is beneficial. The main concern regarding throttling in the context of net neutrality can be found in Recital 15 of the EU's net neutrality regulation. It states that traffic management tools like throttling could be used as a substitute for network capacity expansion (EU regulation 2015/2120). This concern seems to be justified only at first glance. The overall effect of throttling is composed of two opposing effects. In practice, throttling, i.e., limiting the data transmission rate for some content, means that on the one hand, e.g., videos, are displayed to the users with a lower resolution, which is detrimental to users' quality of experience (QoE). On the other hand, the network load per video stream decreases, which in turn has a positive effect on the frequency and strength of congestion events in mobile networks and is thus beneficial for users' QoE. All in all, the impact of throttling on QoE is not obvious and as a consequence the influence on capacity investments is also not obvious, as QoE directly impacts users' willingness to pay and thus the monetization of capacity which in turn is the key driver for capacity investments.

This paper investigates this relation of MNOs' overall capacity investment incentives and throttling by means of a stylized game theoretic model with a monopolistic MNO.

The model shows that the ability of an MNO to throttle enables it to optimize the usage of its mobile network. Surprisingly, this optimization never results in lower incentives to invest in network capacity but rather yields sometimes higher incentives, i.e., a MNO that is prohibited to throttle data streams caused, e.g., by some form of net neutrality regulation will not provide more but sometimes less network capacity to its customers.

The paper is structured as follows. First, a brief overview of the literature is given, dealing primarily with the economic literature on net neutrality. Afterwards a game theoretical model is developed and the results are analyzed and interpreted. Finally, an outlook for future research is given.

2 Related Literature

The paper is mainly connected to the literature on net neutrality. In 2003, a paper of Wu (2003) initialized a debate on the discrimination-free treatment of data in the Internet. The major focus of the economic literature on this topic is about pros and cons to allow paid prioritization of data streams. There are for example studies that deal with the behaviour of a Internet service provider (ISP) which is vertically integrated with a content provider or studies that analyze how paid prioritization influences content providers' incentive to develop innovations (Easley et al., 2018).[1] The papers on net neutrality which are closest to this paper deal, among other, with the impact of paid prioritization on the incentives of an ISP to invest in its infrastructure (e.g., Economides and Hermalin, 2012; Krämer and Wiewiorra, 2012; Choi and Byung-Cheol, 2010). One of the major results of these papers is that paid prioritization yields an additional revenue stream that allows an additional monetization of the mobile network capacity especially if providing more capacity results in an overall increase of content providers' bit rate consumption (see e.g. Economides and

[1] For a detailed literature review on net neutrality see Easley et al. (2018) and Krämer et al. (2013)

Hermalin, 2012; Krämer and Wiewiorra, 2012), i.e., incentives to invest in capacities are lower under net neutrality in such scenarios.[2]

Note, however, that the present paper deals not with paid prioritization but rather with throttling, i.e., a network traffic management tool which comprises no additional revenue stream from content providers. Nevertheless, it will be shown that the ability to throttle can also yield a better monetization of capacity which results in higher incentives to invest in capacity.

3 A Model of Throttling's Impact on Capacity Investment Incentives

3.1 The Model Setup

Consider a market where consumers access the Internet and its content through a monopoly MNO that provides a mobile network with an overall capacity k for a fixed flat rate price p. Without regulation, the MNO is allowed to use throttling, which means that it can limit the bit rate consumption of some content units. The degree of throttling will be denoted as $t \in [0, \infty]$, with a higher t representing higher degrees of throttling and $t = 0$ representing a situation without throttling. The degree of throttling affects the quality of experience (QoE) for consumers and thus their utility from accessing the Internet through the mobile network. The overall impact on QoE can be broken down into two effects.

Firstly, throttling has an impact on the appearance of congestion in peak traffic periods. In particular, throttling lowers the average traffic consumption per content unit since, e.g., each video stream uses a lower bit rate, which, in turn, leads to less frequent and less severe congestion events in the network. Thus, throttling can improve the quality of service (QoS)

[2] This positive effect on capacity investments is countervailed if capacity increases are not connected to more usage on the content provider side. In such cases more capacity reduces the ISPs' ability to discriminate content providers and thus to charge for prioritization (see e.g., Choi and Byung-Cheol, 2010; Baake and Sudaric, 2018).

provided by the network. Note, that QoS is not only influenced by throttling but also by the overall network capacity k, i.e., a higher network capacity as well as a higher degree of throttling is appropriate to prevent congestion. This relation reflects the concern that usage of throttling substitutes capacity expansion. The in-depth analysis, however, will show that it is indeed not a substitutional but rather a complementary relation. In the model QoS is denoted as $Q_S(t,k) \in [0,1]$, where $Q_S = 1$ represents a network without any congestion and $Q_S = 0$ represents a network which allows no data transfers. The positive impacts of throttling degree and capacity on QoS imply $Q_S(t,k) < Q_S(\tilde{t},k)$ for any $t < \tilde{t}$ and $Q_S(t,k) \leq Q_S(t,\tilde{k})$ for any $k < \tilde{k}$.

Secondly, despite the influence of throttling on congestion there is a direct effect of throttling on users' QoE. Throttling affects the average quality of content units not only during peak traffic periods but all the time. That means that with throttling consumers always have to be contented with, e.g., video streams having a lower resolution even though the available network capacity would sometimes allow for higher bit rates. This direct effect of throttling on QoE is denoted as $\Theta_E(t) \in [0,1]$ with $\Theta_E(t) > \Theta_E(\tilde{t})$ for any $t < \tilde{t}$. It is assumed that Θ_E is twice differentiable in t and multiplicatively separable from Q_S.

The influence of changes in QoE comes through two aspects. Firstly, consumers' utility, i.e., willingness to pay per content unit changes. For simplicity it is assumed that willingness to pay per content unit equals QoE and is represented by $Q_S(t,k)\,\Theta_E(t)$. Secondly it is reasonable to assume that QoE influences users' consumption profile. Let $x \in [0,\overline{x}]$ be the amount of content consumed by a consumer. Its content consumption is the higher the less the QoE is affected by throttling or capacity constraints, i.e., $x(t, Q_S(t,k)) > x(\tilde{t}, Q_S(t,k))$ for $t < \tilde{t}$ and $x(t, Q_S(t,k)) \leq x(t, Q_S(t,\tilde{k}))$ for any $k < \tilde{k}$.

In order to limit the complexity of the model, it is assumed that consumers using the mobile network are homogeneous. Consequently, the number of consumers n must assumed to be exogenous. Based on these assumptions, the utility function for each consumer looks as follows:

$$U = x(t, Q_S(t,k))\, Q_S(t,k)\, \Theta_E(t), \tag{1}$$

The fact that there is no consumption if the network is fully throttled implies $x = 0$ for $t = 1$. The same applies to a network which allows no data transfer, i.e., $Q_S = 0$ implies $x = 0$. Finally, to reflect the absence of throttling's and congestion's negative impact it is assumed that $x(0,1) = \overline{x}$. For similar reasons, it will be assumed that $\Theta_E(0) = 1$ in order to address that there is no impact of throttling on QoE without throttling and $\Theta_E(\infty) = 0$ in order to address consumers' inability to access Internet content if all content is fully throttled. The QoS aspect Q_S is modeled as the ratio of mobile network's capacity k and total peak traffic T, i.e.,

$$Q_S = min\left\{\frac{k}{T(t,k)}, 1\right\}, \tag{2}$$

which is a specification referring to Peitz and Schütt (2016). It reflects that if the available network capacity is sufficient to deliver content always as soon as necessary, then QoS is optimal, i.e., congestion is not an issue and $Q_S = 1$. If congestion is an issue, i.e., if the total peak traffic T overreaches the available bandwidth, then QoS is reflected by the probability that a packet is sent in time which is specified as $Q_S = \frac{k}{T(t)} \in [0, 1)$ (Peitz and Schütt, 2016). The total peak traffic itself depends on the number of consumers and their respective data consumption. The MNO can influence this traffic through throttling. The influence of throttling on the data rate consumption per content unit is represented by a function $\Theta_S(t) \in [0, 1]$. This results in

$$T = n\, x(t, Q_S(t,k))\, \Theta_S(t), \tag{3}$$

where $\Theta_S(t)$ is assumed to be decreasing and twice differentiable in t. In other words, more throttling decreases total peak traffic. Further, it is reasonable to assume $\Theta_S(\infty) = 0$ in order to address the absence of traffic in a situation where all content is fully throttled and $\Theta_S(0) = 1$ in order to address that there is no impact of throttling without throttling. Thus, $\overline{T} = n\,\overline{x}$ is the unrestricted peak traffic that occurs if there is no throttling and congestion in the mobile network. Note that consumers do not anticipate their impact on total peak traffic. This assumption is not only in-line with the related literature (Economides and Hermalin, 2015; Choi and Byung-Cheol, 2010) but also with reality where consumers' information regarding their weight on mobile cells' capacity in relevant area is very limited.

The MNO is assumed to incur costs from providing network capacity represented by a quadratic cost function. Other costs are assumed to be zero. Further, the MNO charges an access price p to each consumer. Consequently, the MNO's profit function is

$$\pi = n\,p - k^2. \tag{4}$$

The timing of the model is as follows:

Stage 1: The MNO sets network capacity k, throttling degree t, and access price p.[3]

Stage 2: Users decide to buy or not to buy access to the mobile network and use it depending on the QoE offered.

3.2 Equilibrium Derivation

The game is solved through backwards induction. As consumers are homogeneous, all of them will buy at Stage 2 if and only if $U \geq p$. Anticipating that, it is obvious that the MNO's profit maximizing price at Stage 1 is $p^* = U$. Inserting this result into the profit

[3] By the Envelope Theorem, the results do not change if capacity is chosen in an additional former stage.

function yields in the following maximization problem regarding the choice of t and k:

$$\max_{t,k} \left(\pi = \left(n\, x(t, Q_S(t,k))\, Q_S(t,k)\, \Theta_E(t) - k^2\right)\right) \tag{5}$$

First, note that the inner solution of Equation 5 is $t^* = \arg\max_t \left(k\, \frac{\Theta_E(t)}{\Theta_S(t)}\right)$ and $k^* = \frac{1}{2}\frac{\Theta_E(t)}{\Theta_S(t)}$. Second, note that $k > \overline{T}$ is never a reasonable choice which implies the existence of cases with a corner solution $k^* = \overline{T}$. It is further obvious that in such a case the profit maximizing throttling level is $t^* = 0$, as a higher degree of throttling provides no advantage if $T = k$ but would strictly lower consumer's willingness to pay. In order to specify the appearance of cases where the MNO chooses this corner solution consider that there usually exists a border case $\tilde{T} = \overline{T}$ where the inner solution of Equation 5 with respect to k is equal to the corner solution $k = \overline{T}$. Considering $k^* = \frac{1}{2}\frac{\Theta_E(t^*)}{\Theta_S(t^*)}$ and $t^* = 0$ for $\overline{T} = k$ yields $k^* = 1/2$ for $k^* = \overline{T}$, i.e., the border case is characterized by $\tilde{T} = 1/2$. In summary, the equilibrium choices look as follows:

Lemma 1.

$$p^* = U(t^*, k^*), \tag{6}$$

$$t^* = \begin{cases} \arg\max_t \left(k^*\, \frac{\Theta_E(t)}{\Theta_S(t)}\right) & \textit{if } \overline{T} > 1/2 \\ 0 & \textit{else,} \end{cases} \tag{7}$$

$$k^* = \begin{cases} \frac{1}{2}\frac{\Theta_E(t^*)}{\Theta_S(t^*)} & \textit{if } \overline{T} > 1/2 \\ \overline{T} & \textit{else.} \end{cases} \tag{8}$$

3.3 Interpretation and Results

The intuition regarding t^* is as follows. If congestion is not an issue in the mobile network, i.e., if $\overline{T} \leq k$, there is no reason to lower the QoE by setting $t > 0$, i.e., no throttling is the profit maximizing choice. If congestion is an issue, i.e., if $\overline{T} > k$, then the detrimental influence of throttling on Θ_E must be weighted against its positive influence on QoS, i.e., some positive degree of throttling is profit maximizing as long as $\frac{\Theta_E}{\Theta_S}$ has its global maximum at $t \in (0, \infty)$.[4] This result is summarized in the following proposition:

Proposition 1. *An MNO uses throttling if and only if it can thereby alleviate consumers' disutility from congestion in its mobile network.*

Correspondingly, the profit maximizing capacity is chosen in such a way that either no congestion occurs or that the detrimental effect of congestion is weighted against the capacity costs and the effect of throttling on consumers' willingness to pay. Thereby one can observe that the profit maximizing k for $\overline{T} > 1/2$ is $1/2$ if no throttling is used, i.e., if $t = 0$. Consider further that the MNO chooses t in a way that maximizes $\frac{\Theta_E}{\Theta_S}$. Note that $t = 0$ results in $\frac{\Theta_E}{\Theta_S} = 1$, i.e., an MNO will only choose a throttling degree other than $t = 0$ if it yields $\frac{\Theta_E}{\Theta_S} \geq 1$. Considering this with regard to Equation 8 shows that the ability to throttle never yields lower but rather higher incentives to invest in capacity. In particular, as the unrestricted MNO chooses t in a way that maximizes $\frac{\Theta_E}{\Theta_S}$, its choice will also maximize its incentive to invest in the network capacity.

Proposition 2. *An unrestricted MNO will always choose the throttling level that yields the highest incentives to invest in network capacity.*

Restricting the MNO's choice to use throttling will therefore never lead to an improvement in the overall network capacity, but rather bears the risk of reducing the incentives to invest in network capacity. The intuition behind this at a first glance counter-intuitive result is that each unit of network capacity is used more efficiently through throttling and becomes

[4] Note that $t = \infty$ results in $\overline{T} \leq k$, where $t = \infty$ is weakly dominated by $t = 0$.

therefore more valuable, i.e., throttling increases utility per capacity unit and thus consumers' willingness to pay per capacity unit rises. Consequently, the marginal revenue per capacity unit increases which makes it profitable for the MNO to incur more costs from capacity investments.

4 Conclusion

Summary and Theoretical Contribution: This paper focuses on the MNO's incentives to invest in network capacity, if it is allowed to throttle data streams. By developing a stylized game theoretical model with a MNO operator providing access to the Internet for its costumers, it has been shown that obligations like EU's regulation of net neutrality that prohibit throttling practises do not yield higher incentives for MNOs to invest in mobile network capacity but rather bear the risk to decrease these incentives. This result is caused by the throttling's ability to provide a more efficient usage of network capacity with regards to consumers' overall quality of experience and thus their willingness to pay. Additionally, it has been shown that the MNO uses throttling in order to alleviate consumers' disutility from congestion, i.e., to improve consumers' quality of experience.

Managerial Implications: In terms of managerial implications the model shows that throttling can be an appropriate instrument to improve the QoE of customers and thus their willingness to pay for mobile access. Nonetheless, it is important to note that the detrimental effects of throttling on users' experience from lower quality of, e.g., video content have to be weighted against the beneficial effects through severe and less frequent congestion. Furthermore, MNOs have to consider that throttling gives no reason to reduce network capacity investments but rather to increase them.

Limitations and Future Work: Note, that the model has a limitation in terms of not taking into account competition between MNOs. However, even in a monopoly situation throttling is used by the MNO in order to alleviate consumers' disutility from network congestion, i.e., to make the mobile network more attractive. It is obvious that this cause is even more important under competition. Thus, the motivation of MNOs to use throttling persists if competition is taken into account and consequently the qualitative effects remain the same. Another limitation is the assumption that consumers are homogeneous. This is of special importance for future research that looks on MNOs' ability to use throttling as an instrument for price discrimination. A paper of Zhai et al. (2018) on throttling in the context of cloud computing implies that such a discriminatory intention can be the reason why capacity is artificially scarce by the MNO in order to sufficiently differentiate a lower quality offer for consumers with lower willingness to pay from a higher quality offer for consumers with higher willingness to pay. The same mechanism is identified by Baake and Sudaric (2018) in the context of net neutrality with a focus on paid prioritization. However, Baake and Sudaric (2018) find that the result of Zhai et al. (2018) only holds if capacity is not scarce. In contrast, when capacity investments are sufficiently costly, the ability to discriminate works in favor of capacity investments. This positive effect ones again comes through a better monetization of each capacity unit. One can therefore expect that the ability to discriminate through throttling will usually amplify the effects presented in this paper.

References

Baake, P. and S. Sudaric (2018). Net neutrality, prioritization and the impact of content delivery networks. Collaborative Research Center Transregio 190, Munich, Working Paper No. 102.

Choi, P. J. and K. Byung-Cheol (2010). Net neutrality and investment incentives. *The RAND Journal of Economics 41*(3), 446–471.

Easley, R. F., H. Guo, and J. Krämer (2018). From net neutrality to data neutrality: A techno-economic framework and research agenda. *Information System Research 29*(2), 253–272.

Economides, N. and B. E. Hermalin (2012). The economics of network neutrality. *Rand Journal of Economics 43*(4), 602–629.

Economides, N. and B. E. Hermalin (2015). The strategic use of download limits by a monopoly platform. *Rand Journal of Economics 46*(2), 297–327.

Krämer, J. and L. Wiewiorra (2012). Network neutrality and congestion sensitive content providers: Implications for content variety, broadband investment, and regulation. *Information Systems Research 23*(4), 1303–1321.

Krämer, J., L. Wiewiorra, and C. Weinhardt (2013). Net neutrality: A progress report. *Telecommunications Policy 37*(9), 794–813.

Meyer, D. (2017). Net neutrality: Key zero rating decision made by germany. ZDNet. Retrieved from `https://www.zdnet.com/article/net-neutrality-key-zero-rating-decision-made-by-germany/`, Accessed on 21/11/2018.

Peitz, M. and F. Schütt (2016). Net neutrality and inflation of traffic. *International Journal of Industrial Organization 46*, 16–62.

Qishin, T. (2018). U mobile's giler postpaid and prepaid plans offer unlimited mobile data. thestar.com.my. Retrieved from `https://www.thestar.com.my/tech/tech-news/2018/06/26/umobile-launches-unlimited-mobile-internet-plans/`, Accessed on 21/11/2018.

Saunders, E. (2018). Telkom mobile data – a lot more data behind the scenes. fortressofsolitude.co.za. Retrieved from https://www.fortressofsolitude.co.za/telkom-mobile-data/, Accessed on 21/11/2018.

T-Mobile USA (2018). Cell phone data plans. t-mobile.com. Retrieved from https://www.t-mobile.com/cell-phone-plans, Accessed on 21/11/2018.

Welch, C. (2018). Unlimited data plans are a mess: here's how to pick the best one. The Verge. Retrieved from https://www.theverge.com/2018/6/28/17507322/best-unlimited-data-plans-verizon-att-tmobile-sprint, Accessed on 21/11/2018.

Wu, T. (2003). Network neutrality, broadband discrimination. *J. on Telecomm. & High Tech. L. 2*, 141.

Zhai, Y., M. Stinchcombe, and A. Whinston (2018). Service throttling in cloud computing. Retrieved from https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1350&context=icis2018, Accessed on 21/11/2018.