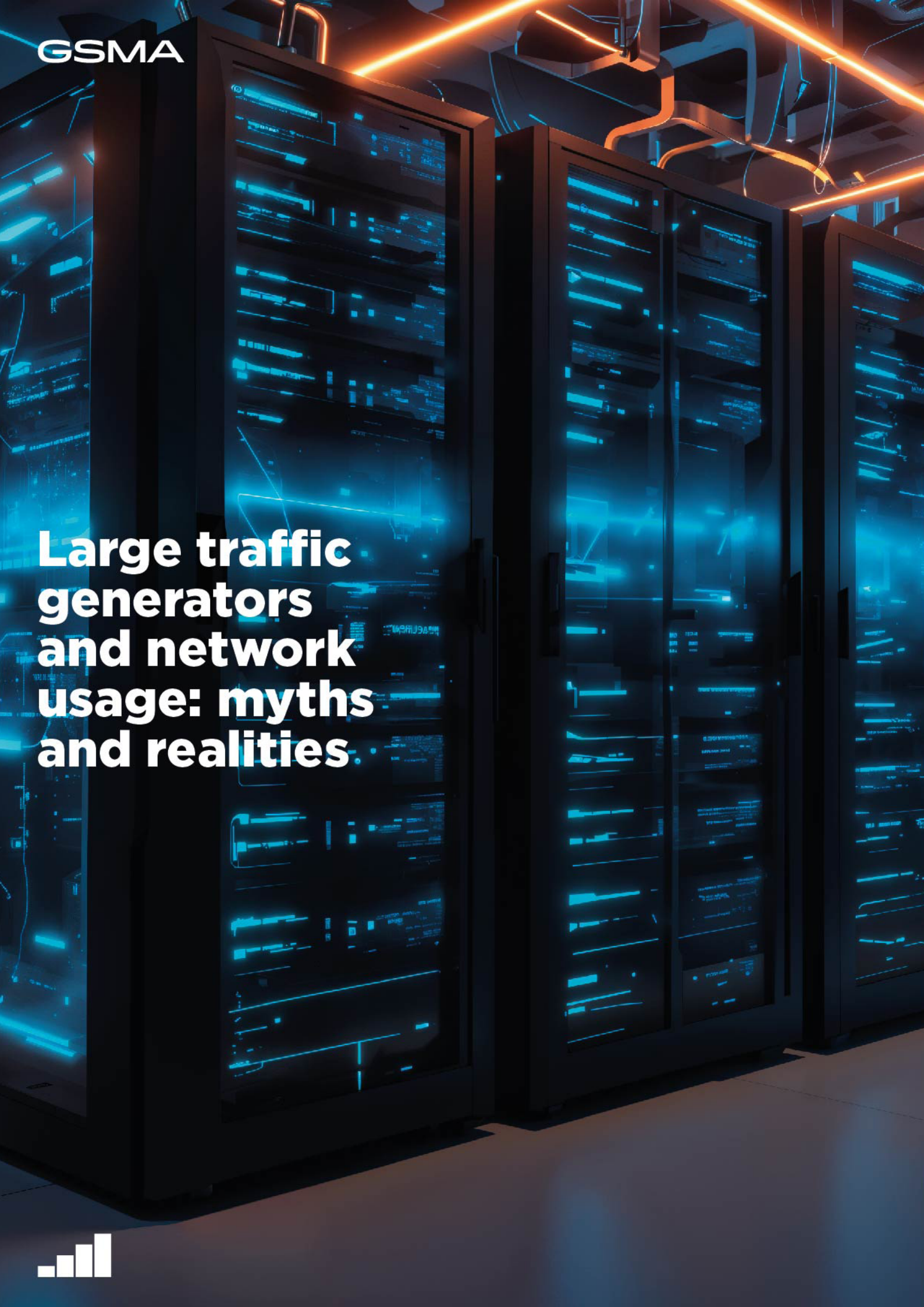
GSMA



Large traffic generators and network usage: myths and realities

GSMA

The GSMA is a global organisation unifying the mobile ecosystem to discover, develop and deliver innovation foundational to positive business environments and societal change. Our vision is to unlock the full power of connectivity so that people, industry, and society thrive. Representing mobile operators and organisations across the mobile ecosystem and adjacent industries, the GSMA delivers for its members across three broad pillars: Connectivity for Good, Industry Services and Solutions, and Outreach. This activity includes advancing policy, tackling today's biggest societal challenges, underpinning the technology and interoperability that make mobile work, and providing the world's largest platform to convene the mobile ecosystem at the MWC and M360 series of events.

GSMA Latin America is the branch of the GSMA in the region. For more information in English, Spanish and Portuguese, please visit www.gsma.com/latinamerica.

Follow the GSMA Latin America on X @GSMALatam and LinkedIn GSMA Latin America.

GSMA Intelligence

GSMA Intelligence is the definitive source of global mobile operator data, analysis and forecasts, and publisher of authoritative industry reports and research. Our data covers every operator group, network and MVNO in every country worldwide — from Afghanistan to Zimbabwe. It is the most accurate and complete set of industry metrics available, comprising tens of millions of individual data points, updated daily.

GSMA Intelligence is relied on by leading operators, vendors, regulators, financial institutions and third-party industry players, to support strategic decision-making and long-term investment planning. The data is used as an industry reference point and is frequently cited by the media and by the industry itself.

Our team of analysts and experts produce regular thought-leading research reports across a range of industry topics.

www.gsmaintelligence.com
info@gsmaintelligence.com

Authors:
Pau Castells, Head of Economic Analysis
Jakub Zagdanski, Senior Economist

Contents

GSMA

1. Executive summary

Digital ecosystem players and policymakers have been engaging in discussions over the viability of proposed payments for network use from large internet traffic generators (LTGs) to network operators and internet service providers (ISPs). Payments have been proposed as a market-based solution that could improve consumer outcomes and help achieve investments, connectivity and digital society goals.

We examine the economic case for a potential market-based solution where LTGs face a price signal for the usage of the public network. By focusing on a general principle rather than a specific scheme design, we review the potential of payments as an instrument to improve economic efficiency, which could translate into improved consumer and societal outcomes, such as greater network quality, increased innovation and a general improvement in the pace of digitalisation and the benefits this brings to society.

Answers to frequently asked questions

We walk through the key questions for and against the existence of payments for network use from LTGs. As these questions have been an object of disagreement, we outline them in non-technical terms and examine them against evidence to support or refute the claims made so far.

We conclude that under the current regulatory framework, the outcome can be suboptimal due to insufficient incentives for LTGs to use networks efficiently. Payments on the basis of network use by LTGs could improve efficiency in the use of networks and increase economic welfare.

The following six key questions confront the myths with realities:

1. Is the amount of network traffic a driver of the network cost? Yes, the amount of traffic drives the network cost. Both the capital expenditure to set up the network and the operating expenditure are driven by the network's design capacity, with increasing capacity directly affecting the network cost. Other factors also drive costs, but traffic volume is a key contributor.

2. Network operators already charge consumers for internet access, so will contributions from internet companies on the basis of traffic delivery amount to double charging for the same service? No. Networks are a two-sided market: consumers use them to reach content, and content providers rely on them to reach consumers. Whether one side or both sides pay for their use depends on various characteristics of each side of the market. Having a contribution from two sides means the cost is shared among both types of network users: consumers and LTGs.

3. What is the advantage of charging LTGs for traffic delivery, compared to just charging consumers for the full amount? Charging on the basis of network use provides incentives necessary to foster network efficiency and avoid the tragedy of the commons. These incentives cannot be solely extended to consumers because they have limited control over consumption and transmission of data. An incentive should be extended to LTGs, which have the ability and the

expertise to manage data flows efficiently, including how their own services generate data traffic. This change could also improve fairness in who bears the network cost: light network users are currently subsidising heavy service and app users.

4. LTGs pay for servers and own content delivery networks, so does this not give them a sufficient incentive to deliver content efficiently to consumers? No. While LTGs have some internal incentives to use parts of the networks efficiently, they are not adequate to enable efficient traffic management. This is because of the cost asymmetry: close to 80% of the total cost of networks is due to access networks, which are financed and managed by ISPs. LTGs undertake only limited investment to optimise data transmission from their cache servers into ISPs' networks and only to the extent it benefits them individually.

5. Why are direct payments so rare today? Currently, regulations significantly constrain network operators' bargaining power, which limits the adoption of network use charges. Examples include the net-neutrality regulations, universal service and quality obligations, and additional rules on peering and interconnection. These contribute to the bargaining power asymmetry, which means that network operators lack the ability to negotiate commensurate agreements that extend incentives to use networks efficiently to LTGs.

6. Would direct payments lead to discrimination on the internet and create a barrier to entry for smaller players? No. Payments from LTGs are not at odds with open internet principles. Compensation for generated traffic can apply equally on a per unit of data traffic basis, rather than to specific types of traffic or specific providers. Such payments could only apply to content providers that achieve large scale and are therefore drivers of network investment requirements, allowing smaller services to test and innovate without an additional transaction burden. Similarly, smaller ISPs could benefit from more equal bargaining power.

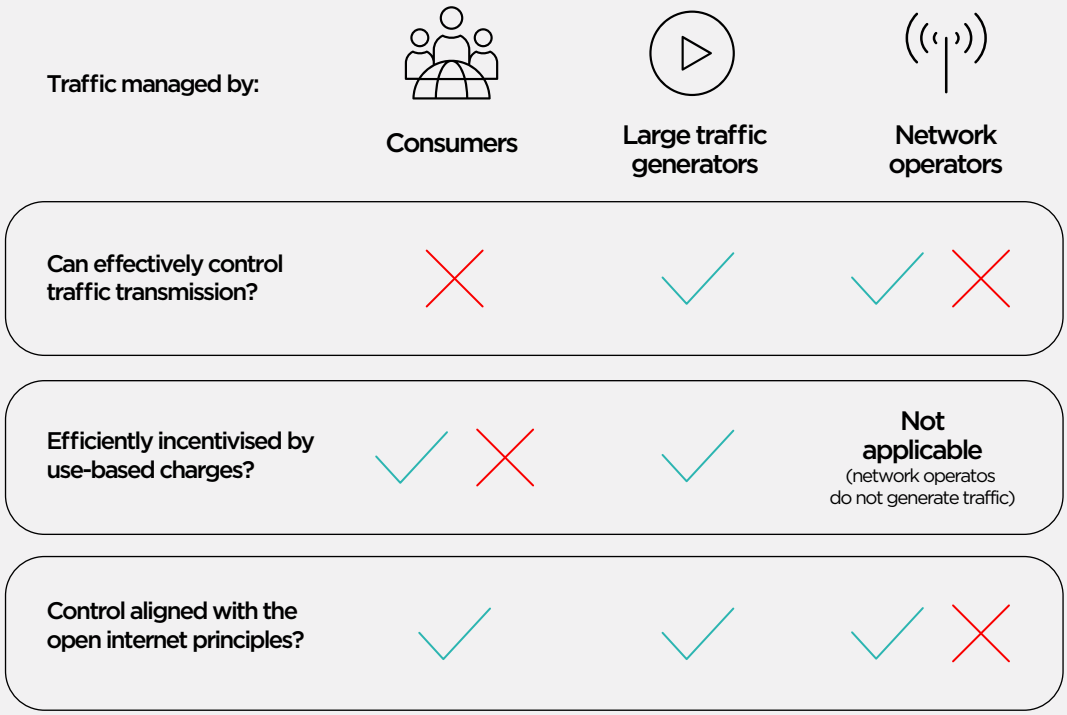
Our conclusions

Taken simultaneously, these answers lead us to conclude that the outcomes in the current framework may be suboptimal and that an additional set of incentives will likely improve economic efficiency and societal outcomes. The effectiveness of incentives depends on which party they are placed on. Consumers do not have sufficient control over how efficiently their request for content will be transmitted. In contrast, LTGs can effectively manage the volume of traffic and use of networks, for example by optimising video quality and data volumes.

A regulatory change that leads to expanded incentives for the LTGs to manage traffic could solve the issues of economic efficiency, as they are better placed to respond to these. Network use payments by LTGs could emerge as a market-based solution, improving investment prospects to deliver the networks of the future.

Figure 1

Who can manage network use the most efficiently?



Source: GSMA Intelligence

2. Introduction

Direct payments are now on the table

The possibility of payments for the use of the public network from large internet traffic generators (LTGs) is now on the table. These payments have been proposed as a market-based solution that could improve consumer outcomes and help achieve connectivity goals.

In this paper, we examine the economic case for a potential market-based solution where LTGs face a price signal to use the public networks run by network operators and the internet service providers (ISPs). By focusing on a general principle rather than a specific scheme design, we review the potential of payments as an instrument to improve economic efficiency, which could translate into improved consumer and societal outcomes, such as greater network quality, increased innovation and a general improvement in the pace of digitalisation and the benefits this brings to society.

How internet apps and services reach consumers today

The prevailing model in which digital infrastructure and services operate today is based on largely separate entities of infrastructure companies (network operators and ISPs) and content and application providers (CAPs).

CAPs rely on public networks run by ISPs, so that consumers can reach their content. At the same time, services and applications on the internet drive consumer demand for connectivity services offered by ISPs. These basic relationships have been in place since the early days of the internet. However, the internet has evolved:

- The internet is a network of connected networks. In the past, ISPs connected their networks relying on settlement-free peering, when growing the network was seen as a win-win situation between ISPs whose traffic was close to symmetrical.
- In the last two decades, a few global CAPs have achieved a very large scale to the point where eight players (Alphabet, Meta, TikTok, Netflix, Amazon, Microsoft, Apple and Disney+) are responsible for nearly 70% of global internet traffic.[1] These, and other large companies in particular countries, are recognised as LTGs.
- Content delivery by LTGs has been optimised by the use of content distribution networks (CDNs). CDNs are cache servers used by LTGs, where content is stored to help deliver quality user experience to their clients. LTGs sought to locate CDNs close to the consumers, often entering agreements with ISPs to establish interconnection between their CDNs and ISPs' access networks. This generally occurs without any ICX or other costs to the LTGs.
- Hence, network investments are currently focused on expanding capacity to accommodate largely asymmetrical traffic from LTGs rather than extending coverage to new users. The cost of meeting the increasing capacity is higher for access networks run by ISPs than for the centralised CDNs run or paid for by LTGs.

1 2024 Global Internet Phenomena Report, Sandvine 2024

This has sometimes resulted in difficulties in finding agreements satisfactory to both ISPs and LTGs. Some LTGs turned to regulators, claiming that ISPs have the power to arbitrarily demand charge for interconnection.[2] At the same time, ISPs have stated the unwillingness to find compromise and the abuse of market power by LTGs in negotiations, alleging that the bargaining position is unfavourably distorted by regulation. In their arguments, LTGs have frequently referred to the need to comply with net-neutrality regulations present in some countries. In general, network operators and ISPs maintain that regulations on how the internet traffic can be handled unintentionally distort bargaining power, requiring the ISPs to treat all inbound traffic in the same way, but putting no similar obligation on the LTGs.

2 "Netflix vs. Comcast 'Net Neutrality' Spat Erupts After Traffic Deal", Time, March 2014

3.
Arguments so far: six key questions

The proponents of direct payments claim that such a solution will lead to better outcomes at an aggregate level, compared to the status quo. These arguments often consider inefficiencies and the inadequacy of incentives to optimise traffic generation at present, the potential free-riding problem due to shared infrastructure use, and the insufficient monetary incentives for ISPs to conduct further investment into networks – all leading to a missed opportunity for improved services and accelerated digitalisation.

The opponents to direct payments have argued that the current outcome is efficient and that any payment for network use would be detrimental to outcomes. They also refer to the existing incentive structures, arguing that these already provide sufficient incentives to use networks efficiently for LTGs, consumers and network operators, thus leading to an efficient outcome.

We note that no robust empirical evidence has examined how direct payments impact market and consumer outcomes so far. The available theoretical examinations have primarily been concerned with specific policy designs, often considering only few of the logical steps in the argument or taking the assumptions underpinning the argument as facts.

To address this gap, this study provides a careful step-by-step assessment of the arguments from an economic standpoint, discussing at each step the validity of the logic and assumptions under the available evidence.

1. Is the amount of network traffic a driver of the network cost?

The myth:

We have seen arguments claiming that the amount of traffic is not a driver of the costs of building and operating the network. These claim that the network cost arising from data traffic generated by LTGs (monetary cost or its environmental carbon cost per GB) is negligible.[3]

Some LTGs note that the cost of interconnection with LTGs' CDNs are low and, therefore, the LTGs do not drive the network cost. Similarly, some claim that the marginal cost per additional unit of data is near zero, given that the infrastructure is already in place and that for much of the time the infrastructure is utilised at below its peak throughput.[4] For fixed networks, they state that the access network is dimensioned around the number of subscribers, not traffic.

In support of some of these arguments, LTGs bring up the fact that the cost per unit of traffic has been decreasing steadily over time, which happened at the same time as the overall network traffic increased.

The reality:

The amount of traffic drives the network cost. Both the capital expenditure to set up the network and the operating expenditure are driven by the network's design capacity, with increasing capacity directly affecting the network cost.

The observed trend in declining cost per unit of traffic is indicative of technological progress. It should not be interpreted as an indicator of the network cost declining as the amount of traffic increases. For example, the following passage suggests that an increase in traffic of over 160% resulted in a network cost increase of only 3%:

> *"In 2018–21, network-related ISP costs increased by 3% in total over three years, whilst network traffic increased by over 160% in that same period, showing that ISP networks can handle significant traffic growth at modest incremental cost."*[5]

This logic is flawed because it compares the network cost across time, and not the cost with and without additional traffic. Without network traffic growth of 160%, the total network cost would have declined over time as a result of technological improvements allowing for transmission of the same amount of data with fewer resources and no further investment.

At all stages in the network investment cycle the marginal (incremental) cost per unit of data is tangible – regulators and researchers acknowledge this in their models of network cost that feature traffic as one of the key determinants of network cost. [6] [7] [8]

3 BEREC preliminary assessment of the underlying assumptions of payments from large CAPs to ISPs, BEREC, 2022

4 Myths Surrounding Network Usage Fees: South Korea, CCIA, 2023

5 The Impact of Tech Companies' Network Investment on the Economics of Broadband ISPs, Analysys Mason, 2022

6 Wholesale Voice Markets Review 2021–26, Ofcom

7 Bottom-up mobile cost model update, Analysys Mason for ICP-ANACOM

8 Estimating Digital Infrastructure Investment Needs to Achieve Universal Broadband, Oughton, Amaglobeli, and Moszoro, 2023

In the most direct terms, data traffic can increase the energy consumption of already installed equipment. It also increases the cost of data centres and network equipment, the capacity of which are scaled up in line with growing needs.

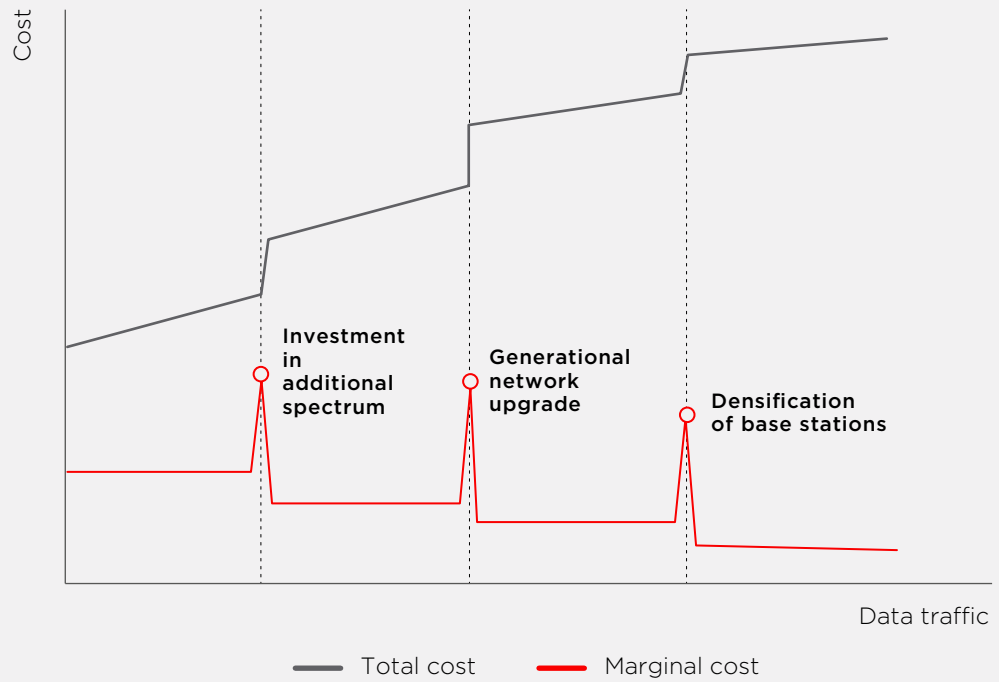
Additionally, meeting the demands of additional traffic can only be achieved via major investments to enhance network throughput. These investments, such as additional base stations, more efficient generations of mobile technology[9] and radio spectrum licences, incur major upfront costs for operators. Hence, while the marginal cost of an additional unit of data may not appear high in the very short term, in the long run the level of traffic determines the entire topology of the network and its cost. Using an example of mobile networks (see Figure 2):

- In low-traffic scenarios, efficiency could mean sparsely distributed base stations with infrequent hardware upgrades.
- In a medium-traffic scenario, efficiency could mean a denser distribution of base stations or investment in more spectrum.
- In a high-traffic scenario, efficiency could mean even denser networks and more frequent hardware upgrades.

Transitioning between each of these topologies incurs major costs. This means that the total cost of operating the network is closely related to data traffic. Given that LTGs account for nearly 70% of global internet traffic, traffic linked to their services is a key driver behind investments in generational upgrades of mobile networks, densification of infrastructure and purchases of additional radio spectrum. Investments are also driven by other factors such as broader IT transformation plans, responding to new cybersecurity challenges or regulatory obligations – but data traffic growth is a key driver.

Figure 2

The relationship between data traffic and network cost



Source: GSMA Intelligence

9 New generations of mobile technology, such as 5G, can carry more data per unit of spectrum i.e. they are more spectrally efficient. For example, 5G can typically transmit 5× more data per unit of spectrum than 3G did.

Data traffic dictates the topology of networks, their density and the frequency of hardware and software updates. Analogous relationships are known from other sectors such as water and energy, where the demand dictates the scale of infrastructure, even though naively it may appear that supplying an additional kilowatt-hour of energy or a unit of water is nearly cost-free. The mobile sector differs only in that mobile traffic per subscriber has increased hundred-fold over the past decade, and similar growth is expected in the near future. This illustrates that growth in demand for data dictates the need for investments in R&D and physical infrastructure and the cost of networks.

In summary, network cost is traffic-dependent to a significant degree. For network use to be efficient, incentives need to be introduced. These incentives need to be set at such a level that internalises the cost that data generates throughout the network.

The importance of an efficient solution to emerge is recognised in digital transformation plans for the networks of the future.[10] [11] [12] These strategies aim to eliminate the connectivity gap and prepare the infrastructure for continued growth in demand for data, both of which will ultimately benefit over-the-top service providers by enlarging the market, enabling support for new revenue streams and resulting in a better online experience due to decongestion for individual users as well as society as a whole.

Some have attempted to play down the importance of addressing the inefficiencies by claiming there is a limited need for further investment in network infrastructure and thus a limited potential benefit of an intervention. This argument primarily rests on two points:

- That the marginal cost of an additional unit of data is negligible once the initial investment in infrastructure is finalised.[13]
- That the future growth in data will be limited and does not warrant further investment into networks.

With regard to the former, data is a direct driver of network cost, which is especially evident in the long term. Upfront network investment is dictated by the demand for data and connection quality – these are what matter to consumers rather than the label on the latest technology, such as fibre or 5G. Hence, demand drives investment, such as generational upgrades (e.g. laying fibre-optic cable or upgrading radio towers). Overlapping technology generations and geographically staggered deployments mean that investments are undertaken continuously. It is untrue that investment will cease once fibre networks and 5G have been rolled out.

With regard to future data consumption, forecasts show that growth will continue. Future data traffic growth will be driven both by the continued improvement in the quality of video transitioning from SD to HD, 4K and 8K, as well as new applications and services relying on AI, AR and VR.

Sometimes, it is cited that traffic growth rates are slowing down. However, this is an effect of increasing base levels. In absolute terms (e.g. GB per connection per month), it is anticipated that the growth in the next seven years will be several times greater compared to the last seven years (Figure 3). Given that LTGs account for a nearly 70% share of total internet traffic, their upgraded services and applications will remain a core and significant driver of required investments in networks and their costs. This growth, along with the next wave of AI and VR use cases, will only be possible with further investment in networks. A potential improvement in efficiency of network use will have a major impact on the pace of digitalisation and the benefits it brings to society.

10 White Paper - How to master Europe's digital infrastructure needs?, European Commission, 2024

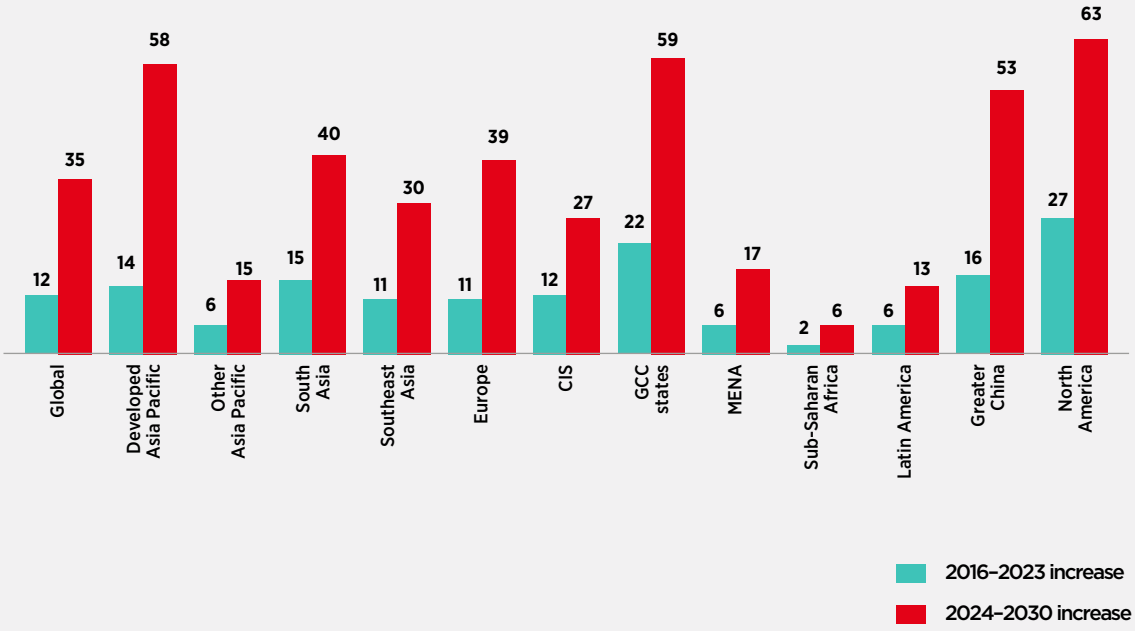
11 1. Connectivity - building world-class digital infrastructure for the UK, DCMS, 2023

12 Brazilian Digital Transformation Strategy, Department of Digital Transformation Policy, 2018

13 BEREC preliminary assessment of the underlying assumptions of payments from large CAPs to ISPs, BEREC, 2022

Figure 3

Forecast of growth in mobile traffic



Source: GSMA Intelligence

2. Network operators already charge consumers for internet access, so will contributions from internet companies on the basis of traffic delivery amount to double charging for the same service?

The myth:

LTGs argue that ISPs have already charged their subscribers for access to the internet and, therefore, LTGs should not contribute further. According to LTGs, any payments from them would amount to double-charging for the service that consumers have already paid for.[14]

The reality:

Networks are a two-sided market: consumers use them to reach content, and content providers rely on them to reach consumers. Whether one side or both sides pay for their use depends on various characteristics of each side of the market.

Networks are an example of a two-sided market, where consumers and LTGs rely on the network to reach each other.[15] Two-sided markets are particularly common in the digital ecosystem, but they can be found everywhere, such as the following examples:

- **The credit card market:** Consumers derive value from being able to pay with their credit card and the sellers can access additional customers by accepting payments underwritten by the credit card companies.
- **The publishing industry:** Readers gain access to content and classifieds, and the advertisers can reach potential customers with their offers.
- **The internet platforms run by LTGs:** Users access content or services, while the platform serves content from producers and advertisers. Examples include platforms such as Facebook, advertising on Google, ride-booking services such as Uber or e-commerce services such as Amazon and eBay. All these platforms are a meeting place for consumers and advertisers.

In a two-sided market, whether one or both sides pay for use depends on the market characteristics. There are examples of each type: publishers of classifieds can offer free access to the readers and charge the advertisers; alternatively, some publications do not charge for posting adverts but charge for access. In simplified economic terms, who the intermediary (the publisher or a network operator) chooses to charge depends on the relative value an additional consumer or an additional content or service provider brings to the pool of existing readers and advertisers.

In fact, LTGs often rely on double-sided charging structures, charging the consumers directly (streaming services) or indirectly (advertising), or both at the same time, as well as charging the content providers/ advertisers (app stores and online auctions).

14 A cooperative approach to content delivery, Netflix 2021

15 Two-sided markets: a progress report, Rochet & Tirole, 2006

The nature of networks means that they connect both sides of the market, allowing consumers to access content and allowing content distributors to access audiences. Operators in this market type can have various approaches to charging one or both sides simultaneously. This does not mean double-charging for a product that has been paid for already.

Double-sided charging structures can evolve over time and a good example of this is mobile voice calls. Originally, both the caller and receiver of a call paid for the service but this later evolved in most markets to a regime in which only the party initiating the call pays, although the double-sided charging scheme still exists for some roaming services.

3. What is the advantage of charging LTGs for traffic delivery, compared to just charging consumers for the full amount?

The myth:

Some have argued that as consumers initiate the traffic, ISPs should charge them in line with the generated traffic.[16] This would mean that it is up to the ISPs and network companies to make sure that they charge customers accordingly to recover the cost and provide the necessary incentives for efficient use.[17]

The reality:

Charging on the basis of network use provides incentives necessary to avoid free riding and the tragedy of the commons. These incentives cannot be solely extended to consumers because they have limited control over consumption and transmission of data. An incentive should be extended to LTGs, which have the ability and the expertise to manage data flows efficiently, including how their own services generate data traffic.

There are limitations to how much ISPs can rely on consumer tariffs to promote efficient network use. There are several reasons for this:

- Consumers have limited control over how much data they use and how efficiently their requests for content are fulfilled.[18] While some tools are available to manage data consumption, these involve a high learning and monitoring cost. Aggregated over billions of consumers, continuous monitoring of data consumption for individual apps and services across multiple devices and configuring settings would amount to a huge time and effort burden. This makes the direct incentive in the form of use-based tariffs inefficient.
- Business models of CAPs rely on cross-promotion of different services to maximise the amount of content viewed and the size of the network, to grow the audiences for advertised content. This means that a sizeable portion of traffic, such as ads or prefetched video, is unsolicited. Studies show that up to 29% of total data traffic could be unsolicited for some of the leading social media platforms.[19] Furthermore, optimising delivery of video traffic has the potential to reduce overall traffic by 15–25% while maintaining the same user experience.[20] These statistics illustrate that far more can be achieved to optimise traffic beyond what is currently being achieved with inefficient incentives only for the consumers.
- Consumers can be highly loss averse: facing use-based tariffs, they would seek such a tariff that minimises or eliminates the possibility of an extremely high bill linked to unplanned use, for example as a result of accidental use, or malicious software.[21] [22] In practice, these consumer

16 Regulatory Myopia and the Fair Share of Network Costs: Learning from Net Neutrality's Mistakes, Colangelo, 2023
17 BEREC preliminary assessment of the underlying assumptions of payments from large CAPs to ISPs, BEREC, 2022
18 In Pursuit of Sustainable Mobile Policy: A Study of Consumer Tariff Preferences under Uncertainty, Han, Yun and Yeo, 2020
19 Characterisation of Unsolicited Traffic Advertisements in Mobile Devices, Silva et al., 2020
20 Mobile Investment Gaps: Caribbean islands (annex), GSMA Intelligence, 2024
21 Uncertain Demand, Consumer Loss Aversion, and Flat-Rate Tariffs, Herweg and Mierendorff, 2013
22 Is Having an Expert "Friend" Enough? An Analysis of Consumer Switching Behavior in Mobile Telephony, Genakos, Roumanias and Valletti, 2015

preferences mean that operators typically offer flat-rate tariffs with unlimited data and tiered speeds (fixed connections), or flat-rate tariffs and data bundles purchased in advance to avoid unpredictable charges (mobile). Flat-rate tariffs can help, although they do not fully remove consumers' loss aversion. Additionally, they mean that low-volume users cross-subsidise high-volume users, which can contribute to the digital divide.

Limited control over data use and loss aversion together mean that distributing the incentive to use data efficiently solely across consumers may be suboptimal and will not provide sufficient incentives to optimise network use.

LTGs are fully aware of the impact of consumers' risk aversion when facing use-based tariffs. LTGs commonly rely on subscription-based flat-rate charging mode for services such as video and music streaming.

In summary, bounded rationality means that consumers alone cannot efficiently respond to use-based tariffs and that alternative incentive structures should be considered to promote efficient network use. A more efficient outcome could be achieved if the incentive structure is extended to LTGs via payment. Charges on the basis of LTGs' traffic can provide an improvement because of the characteristics of LTGs in comparison to consumers:

- LTGs have the expertise and resources to monitor, control and optimise how much and when they generate traffic on the networks. This includes determining codecs, configuring prefetch and bitrates and optimising the quality of video (to a higher or lower degree) to the characteristics of the end-user device.
- LTGs have more risk-neutral preferences, with more predictable aggregate traffic patterns and use averaged over millions of users. This means that charges on the basis of data flows can be better tolerated by LTGs.
- At the same time, volume-based payments from LTGs preserve the existing incentives for ISPs and network operators to build and operate networks efficiently.

A price incentive will internalise the cost additional traffic generates throughout the networks and increase the efficiency of how requests for content are served. This efficiency means that networks could be decongested from unwanted traffic (ads, auto-play) when the total value of this traffic for the sending and receiving party is below the cost it generates throughout the network. This can lead not only to economic benefits due to efficiency in consumption of data and removal of a negative externality, but also environmental benefits in the form of more efficient resource use.

4. LTGs pay for servers and own content delivery networks, so does this not give them a sufficient incentive to deliver content efficiently to consumers?

The myth:

It has been claimed that LTGs are already contributing to the infrastructure to the extent that fully internalises and offsets the cost of their traffic[23] by investing in subsea cables, data centres and CDNs, which are cache servers storing content such as videos. Therefore, LTGs argue that their investments take the load off ISPs' networks and that they are already sharing the burden of traffic costs.[24] [25]

The reality:

While LTGs have some internal incentives to use parts of the networks efficiently, they are not adequate to enable efficient traffic management. This is because of the cost asymmetry: close to 80% of the total cost of networks is due to access networks financed and managed by ISPs. LTGs undertake only limited investment to optimise data transmission from their cache servers into ISPs' networks.

LTGs have argued that CDNs and additional network infrastructure they fund provide various cost savings to ISPs. Firstly, these savings come as a result of on-net location of CDNs, which reduces potential transit charges. Secondly, multiple CDNs on the host ISP's network could be accessed from a nearby server, rather than having to travel across the ISP's core network. Simultaneously, LTGs have been optimising delivery by developing data-saving video encoding.

While LTGs are investing in infrastructure or data-saving, this only pertains to parts of the infrastructure supporting and optimising delivery of their content from their cache servers to the ISPs' core network. Aside from looking to improve the quality of video experienced by their users, these investments are also often linked to other revenue streams for CAPs such as cloud computing. These investments are logically only carried out to the extent that they minimise costs or maximise profits for CAPs, with limited or no consideration of any implications for ISPs' costs.

For example, international connectivity costs for ISPs are directly linked to the CDNs installed, but since LTGs do not face these costs, CDN investments will only be carried out to the extent that they minimise costs for LTGs, not ISPs. Similarly, smaller and rural ISPs often will have no or a very reduced number of CDNs connected to their networks.

As shown in Figure 4 , ISPs operate a highly decentralised network that physically extends to each local area being served. Decentralisation of access networks means a much greater cost. It is estimated that access network accounts for over 80% of the total fixed network cost.[26] In contrast, LTGs' network

23 Internet Society's Submission to the European Commission's Exploratory Consultation on "The future of the electronic communications sector and its infrastructure", The Internet Society, 2023

24 A cooperative approach to content delivery, Netflix, 2021

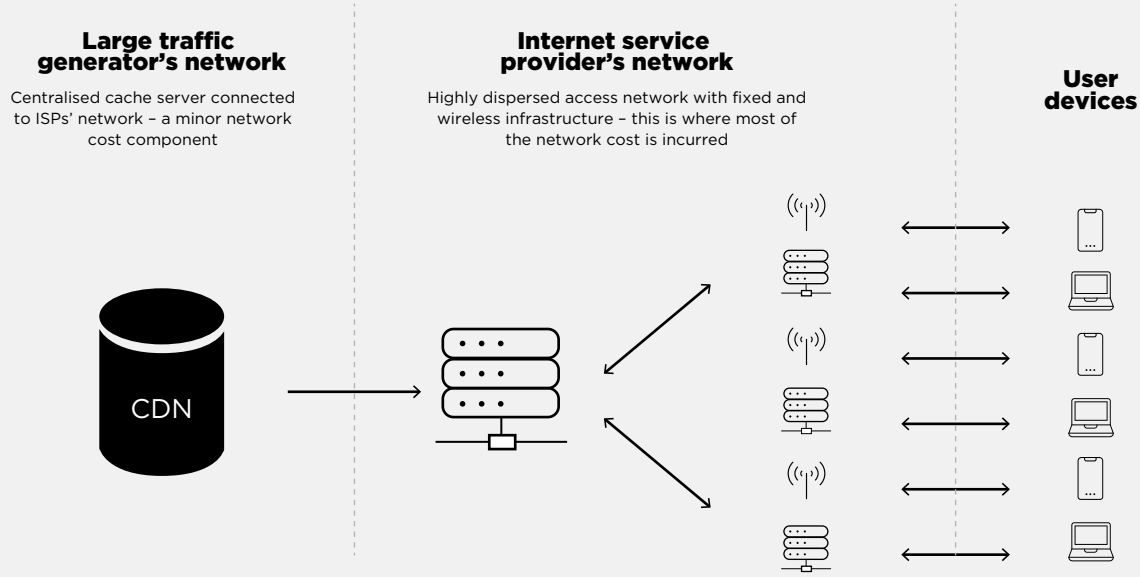
25 "Network Fee Proposals Are Based on a False Premise", Meta, March 2023

26 Net neutrality review, Consultation Annexes 5 to 10, Ofcom, 2022

connections are highly centralised CDNs. For example, Google operates 100 CDNs located in 30 countries, serving as access points to their entire range of services delivered globally to billions of devices via access networks run by ISPs in all countries of the world.

Figure 4

Differences in network architecture mean that ISPs' networks account for the vast majority of data transmission networks



Source: GSMA Intelligence

This cost asymmetry means that the investments by LTGs will not be commensurate to the cost that the traffic generates throughout the entire network. Their investments will not internalise the cost impact located on the ISP's network, in particular the access network. While the presence of LTGs' CDNs offers a relief on some costs, it is not commensurate to the overall contribution of the data traffic they generate to the total network cost.

Investment by LTGs is only undertaken to the extent that it optimises their own content delivery where it is profitable to do so. Over 90% of LTGs' investment is directed into storage, rather than the transport and delivery network.[27] Estimates show that in 2022, LTGs invested approximately $7.6 billion in transport and delivery functions of networks. In contrast, it is estimated that the total capital investment of network operators stood at $200 billion in 2022, with the combined capital and operating expenditure reaching close to $1 trillion in 2021.[28] [29]

In summary, even though LTGs invest in some infrastructure or data-saving functions, it does not follow that the amount of investment is sufficient to internalise the cost of data they generate on telecoms networks. As long as LTGs do not face a price signal that is commensurate to the cost of data traffic they generate, the incentive for LTGs to invest in throughput or optimise data traffic use will be insufficient to deliver efficient outcomes.

27 The Impact of Tech Companies' Network Investment on the Economics of Broadband ISPs, Analysys Mason 2022

28 ibid.

29 GSMA Intelligence Data Portal, 2024

5. Why are direct payments so rare today?

The myth:

Some analyses point to markets as a source of the prevailing model in which consumers pay the network operators' costs through their subscriptions, while content providers do not directly pay for traffic. They claim that the internet has evolved towards this arrangement as an efficient, market-driven outcome, without a need for regulatory change.[30]

The reality:

Currently, regulations and obligations significantly constrain network operators' bargaining power and limit the adoption of network use charges. Examples include the net-neutrality regulations, universal service and quality obligations, and additional rules on peering and interconnection. These contribute to the bargaining power asymmetry, which means that network operators lack the ability to negotiate commensurate agreements that extend incentives to use networks efficiently to LTGs. This bargaining power asymmetry means that network operators lack the ability to extend incentives to use networks efficiently to LTGs.

The historical model of charging consumers and settlement-free peering was designed when traffic was exchanged relatively symmetrically between ISPs. When this was not the case, the asymmetry was settled using fees.

Since then, the internet's architecture has evolved towards ISPs connecting with LTG-run CDNs, where currently nearly 70% of global internet traffic originates.[31] This has meant that traffic largely enters ISPs' networks from CDNs and generates cost that largely falls onto ISPs that run access networks. With this change, network operators began seeking ways to cover that cost.

Various obstacles have contributed to the limited extent to which ISPs and LTGs shifted to commercial, paid network-usage agreements, including free riding and imbalanced bargaining power:[32]

- **Free riding** occurs because even if only some LTGs paid for network use, all of them would benefit from improved infrastructure. Any improvement in the network infrastructure as a result of potential payments or investment will benefit all traffic generators, regardless of whether they have contributed or not.
- **Imbalanced bargaining power** exists because regulation and obligations limit the ways in which ISPs can manage traffic and respond to LTGs' increased volumes of data traffic. ISPs are subject to a number of regulations. These include universal access obligations, rollout and coverage obligations attached to operating and spectrum licences, price controls, quality-of-service regulations and net-neutrality regulations.[33] [34] These have been introduced with the intention of enabling high-quality access in areas or market segments where there is limited commercial attractiveness, preventing potential discrimination of services and maintaining low barriers to entry for new services. However, the same rules do not apply to LTGs or cannot be

30 Internet Society's Submission to the European Commission's Exploratory Consultation on "The future of the electronic communications sector and its infrastructure", The Internet Society, 2023

31 "The Internet is Closer to Home Than You Think", Cisco, June 2017

32 Another Look at the Debate on the "Fair Share" Proposal, Condorelli et al. 2023

33 Subject to price control and type of regulation, ITU Data Hub, 2024 34 UAS Policy, ITU Data Hub, 2024

enforced due to a lack of local presence. As a result, these regulatory asymmetries become a de facto barrier for commercial agreements between parties to take place.

Hence, while aiming for desirable outcomes, regulations on network operators have an unintended restrictive impact on the extent to which networks and traffic can be managed by operators. Some of these regulations have been shown to negatively impact content innovation, investment in networks and the total economic welfare.[35] Importantly, regulations create distorted bargaining power for negotiations between network operators and LTGs on network use arrangements. A few examples of this are listed below:

- **Network operators are required to treat all incoming traffic with equal priority, regardless of its type or the sending party. However, LTGs do not face such obligations, allowing them to differentiate between ISPs and how they send traffic.** In light of the current regulation, LTGs can choose to route their traffic via third parties to any given ISP without regard for the quality of service for the consumers on a specific ISP's network. This leads to an asymmetrically higher cost in terms of quality loss to an ISP if an LTG decides to walk away from negotiations for direct paid interconnection and to route traffic inefficiently. This has, in some instances, led to deteriorated quality of service.[36]
- **Re-routing of services can lead to congestion in certain parts of the network, affecting all services and not just the service specific to an LTG.** This can occur if the routing of LTGs' traffic transits a bottleneck interconnection, for example between two ISPs. The possibility of an LTG walking away from negotiations and instead relying on transit leaves the possibility of creating bottlenecks for which an ISP cannot prepare. This creates a risky, high-cost scenario for ISPs, which are often subject to quality obligations for other traffic as part of commercial agreements or regulatory requirements.
- **Universal service and quality obligations can increase costs for ISPs, as they must expand infrastructure and maintain undifferentiated services in areas where it would not have been commercially viable to do so.** This weakens the financial position of ISPs and their bargaining power against content providers. For example, regulators test service quality based on accessibility of LTGs' services and applications to determine whether the operator is meeting the service-quality obligations.[37] If these are not met due to the lack of agreement and arbitrary routing selected by an LTG, the ISPs are ultimately responsible and liable to the regulator. At the same time, LTGs' position is strengthened by the same regulation.[38] Due to quality and coverage obligations ISPs maintain service quality across their entire network, increasing the size of the potential customer base for LTGs where otherwise it would have been commercially unviable.

In summary, regulation and obligations on network operators create distorted bargaining power and restricts the ways in which they can manage traffic or incentivise LTGs to manage traffic. The distorted bargaining power has meant that even though the internet model has evolved towards a new architecture, market-based agreements and direct payments could not be agreed to align with the new business models. In some instances, regulators have taken a pragmatic approach to enforcement, allowing reasonable network management and exceptions to prioritise latency-sensitive traffic or filter traffic linked to cybercrime.[39] However, these exceptions are insufficient to balance the bargaining power, considering the remaining issues around the disparate impacts of traffic routing and placing the quality-of-service obligations solely on ISPs.

Balancing the negotiating power could facilitate wider adoption of market-based agreements. These agreements could introduce an incentive to use networks efficiently and eliminate the tragedy of the commons that results in suboptimal network quality.

35 Net neutrality and high-speed broadband networks: evidence from OECD countries, Briglauer et al., 2022

36 "Facebook sanctioned for server re-routing that led to user access slowdown", Lexology, March 2018

37 Tender for the provision of mobile quality of service (qos) benchmarking services, Malta Communications Authority, 2024

38 Some LTGs even publish comparisons of ISPs' network quality for their users. For example, see Netflix

39 A survey of Network Neutrality regulations worldwide, Garrett et al., 2022

6. Would direct payments lead to discrimination on the internet and create a barrier to entry for smaller players?

The myth:

It has been claimed that charges on LTGs contravene net-neutrality regulation and will result in unfair discrimination of online services.[40] According to these arguments, requiring LTGs to make monetary contributions will result in substantial transaction costs.[41] If true, this could lead to a decrease in content diversity and become a barrier to entry for smaller ventures due to asymmetry of bargaining power and transaction costs.

It has also been argued that payments for network use from LTGs could lead to ISPs having a monopoly in data traffic termination in a similar manner to the traditional telephony termination monopoly.

The reality:

Payments from LTGs are not at odds with the principles of the open internet. Compensation for generated traffic can apply equally on a per unit of data traffic basis, rather than to specific types of traffic or specific providers.

Direct payments can be compatible with the principles of non-discrimination and openness of the internet. In the context of some of the established principles, a payment system will not be discriminatory if it applies equally to all parties on the basis of the same rules. If these rules apply equally regardless of traffic type or entity type, there is no risk that the openness of the internet would diminish or that certain players will enjoy an advantage over others.

We acknowledge that direct payments could have a disparately high impact on smaller CAPs due to the transaction costs involved, including negotiation and monitoring costs as well as administrative burden. However, any payments could be applicable to only LTGs to minimise the disparate burden on smaller players and preserve the dynamics on the internet. This principle of exemptions for smaller and medium enterprises is already enshrined in regulations worldwide, such as simplified tax reporting, exemptions from some data protection regulations or exclusions from CO_2 cap-and-trade systems for smaller carbon emitters.

Such an exemption from network use charges for all but a few of the largest traffic generators will also ensure that innovators and new entrants will not face an additional barrier to test products and services. Fees could be charged only when entrants achieve the scale of use and traffic that qualifies the service as a LTG. A design with size-based exemptions can ensure no barriers to innovation, preserving the spirt of the internet that made it so successful.

It has also been argued that payments for network use from LTGs could lead to ISPs having a monopoly in data traffic termination, in a similar manner to the traditional telephony termination monopoly.[42]

40 Ten Compelling Reasons to REJECT Cost-Sharing, Internet Society Brasil, n.d.

41 Proposals for a levy on online content application providers to fund network operators, Oxera, 2023

42 BEREC preliminary assessment of the underlying assumptions of payments from large CAPs to ISPs, BEREC, 2022

However, the market power of a monopoly in telephony is not equivalent to the competition dynamics observed in infrastructure-based competition markets in mobile communications.

Similar to how operators compete for and price services for consumers on one side of the market, they would compete for and charge competitively for LTGs on the other side of the market. Competition authorities will continue to oversee pricing and other market outcomes to ensure competition dynamics in mobile and telecommunications markets are beneficial for consumer welfare.

In practice, improved efficiency in network use could benefit innovation, with additional indirect effects due to:

- improved service quality, as the incentive to conserve data could remove congestion from networks
- adequate levels of investment in next-generation network technologies, as payments for data transmission would lead to adequate incentives for the operators.

Finally, transaction costs should only be evaluated in comparison to their present levels given the market failure and lack of regulatory clarity. These are not insignificant, as specific terms are negotiated for each interconnection arrangement while facing a lack of regulatory clarity. In some instances, the disagreements and lack of regulatory clarity have led to costly litigation.[43] [44] Recent cases include the dispute between Meta and Deutsche Telekom in Germany,[45] where the failure to agree on whether direct payment is liable has been brought by different interpretations of regulations. A similar dispute originated in South Korea, where Netflix initially sought to confirm that it cannot be held liable for any direct payment related to network use.[46]

43 Rationales for and Against FCC Involvement in Resolving Internet Service Provider Interconnection Disputes, Frieden, 2011

44 Verizon v. FCC, Court of Appeals, Dist. of Columbia Circuit, 2014

45 See here for Deutsche Telekom's perspective on the case and here for Meta's perspective

46 "Korean court ruling over a network usage fee dispute between Netflix and SK Broadband", Chambers & Partners, July 2021


GSMA

4.
Conclusion

In the preceding chapter, we have evaluated the arguments and established key facts on network cost drivers, market characteristics and the distribution of incentives among consumers, LTGs and ISPs. Taken simultaneously, these lead us to conclude that the outcomes in the current framework may be suboptimal and that an additional set of incentives will likely improve economic efficiency.

The effectiveness of incentives depends on which party they are placed on (see Figure 5). Consumers do not have sufficient control over how efficiently their request for content will be transmitted. In contrast, LTGs can effectively manage the volume of traffic and use of networks, for example by optimising trade-offs such as video quality and data volume. LTGs possess the necessary expertise and knowledge of users' habits, preferences and devices. Network operators have only a limited ability to manage traffic efficiently (e.g. by prioritising time-sensitive data such as calls during congestion) and do not have the technical ability or user-specific information to conduct traffic optimisation with the same efficiency as LTGs.

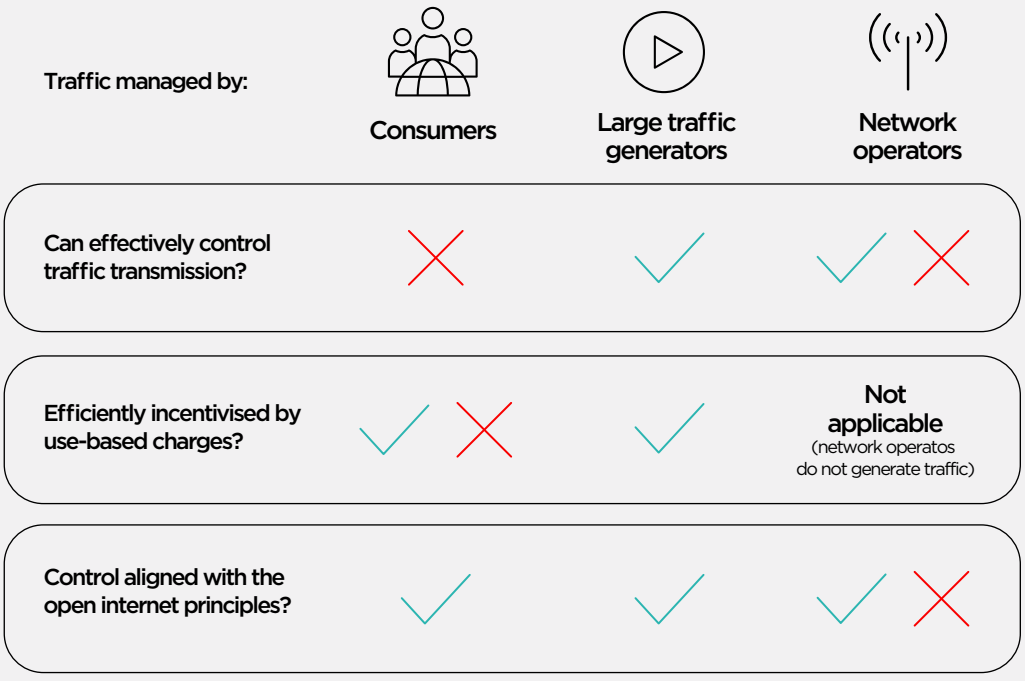
Similarly, consumers can only be partially incentivised by use-based tariffs, as the cost of monitoring and managing data consumption is high for individuals. Consumers are also risk averse, and many could choose to remain unconnected when faced with use-based tariffs while lacking control over data use. In contrast, LTGs have the expertise and resources to monitor data generation and optimise its use in relation to charges.

Lastly, control over consumption of data by consumers or how own data is transmitted by LTGs is aligned with strict net-neutrality rules. Management of own traffic by LTGs and consumers poses no concern over whether this impairs the level of competition since only own traffic is managed.

A regulatory change that leads to expanded incentives for the LTGs to manage traffic could solve the issues of economic efficiency, as they are better placed to respond to these. Network use payments by LTGs could emerge as a market-based solution, improving investment prospects to deliver networks of the future.

Figure 5

Who can manage network use the most efficiently?



Source: GSMA Intelligence

GSMA Head Office
1 Angel Lane,
London,
EC4R 3AB,
UK
Tel: +44 (0) 20 7356 0600