Sustained Throughput Requirements for Future Residential Broadband Service

Traffic Model for Bandwidth Estimates

A Technical Paper prepared for SCTE/ISBE by

Jeroen Wellen
Senior Member of Technical Staff
Bell Labs Consulting/Nokia
Antareslaan 1, Hoofddorp
The Netherlands
Jeroen.Wellen@bell-labs-consulting.com

Prudence Kapauan
Distinguished Member of Technical Staff
Bell Labs Consulting/Nokia
1960 Lucent Lane 9D-117
Naperville, IL 60563
Prudence.Kapauan@bell-labs-consulting.com

Amit Mukhopadhyay
Partner
Bell Labs Consulting/Nokia
600 Mountain Avenue
Murray Hill, NJ 07974
Amit.Mukhopadhyay@bell-labs-consulting.com

Table of Contents

List of Figures

List of Tables

Introduction

The access network is the costliest investment for any operator and it is also the hardest to evolve to next generation technologies. Incumbent operators try to prolong their investment with minimal alterations to the “last mile” connection while new operators try to make their investments future proof as far into the time horizon as practical. While Fiber to the Home (FTTH) is the unquestioned leader from a performance perspective, the economics of the solution works out only in select morphologies and market conditions. Access technologies like Hybrid Fiber Coax (HFC) and Digital Subscriber Line (DSL) comprise the majority of the residential broadband connections today and Fixed Wireless Access (FWA) is also gaining momentum in certain markets. The challenge for DSL is that it can match FTTH performance for only limited loop lengths. HFC is largely constrained by the fact that it operates in a shared medium environment. FWA shares both the limitations – distance as well as shared medium.

The key is to balance the market-specific technology needs and the business justification of the capital and operational expenses. In other words, a technology that may be suitable for delivering the services to a particular geographic and demographic morphology, may be completely unsuitable for delivering a similar service to another morphology. Access networks are considered long term investments and operators make decisions based on revenue potentials and thus Return on Investment (RoI) over a relatively longer period of time is generally acceptable.

Choosing the right Access Network technology at the right moment is one of the most challenging decisions faced by Network Operators. How long the installed base can support the ever-growing demand of applications, especially video, and what technology to install or upgrade to, and when, are the key questions to be answered. HFC and FWA are particularly attractive technologies since they connect multiple homes to a single Access Node (a fiber node in HFC and a cell site for FWA) through a shared medium, allowing to benefit from stochastic multiplexing while minimizing drop costs. But the advantages of these technologies depend on how many homes can be connected per node. And, to assess the techno-economics of the solution, proper estimation of the expected traffic is needed – at present and during the foreseen lifespan of the investment.

This paper presents a model to estimate the 10 years maximum sustained throughput requirements for residential broadband services. The model is used to estimate the maximum throughputs during busy hour and to assess how upper throughput percentiles compare to the mean expected value. In addition, the impact of other effects such as disruptive video applications and loads during special events are investigated. The main purpose of the paper is to provide guidance to operators investing in either upgrading their current network or investing in new network technology. While multi-gigabit access connections often catch the headlines, that is not the sole criteria around which access networks need to be designed. It is imperative that shared medium technologies provide such peak speed connections to support demanding applications. In many circumstances, it is equally important to provide high quality sustained throughput connections at considerably lower speeds. The service targets and the choice of technologies are driven by market economics.

Study

1. Traffic Model

Although traffic demand is basically no more than just the downstream and upstream information rate, it can be viewed at various time scales for various objectives. For the purpose of network design, peak or line rates will determine the maximum distance or range of a technology and thus, by observing the household density, the required number of Access Nodes to cover a geographic area. But to assess equipment requirements per distribution point, or how many radio channels, cable modem terminations or uplink ports are needed at each Access Node, the Maximum Sustained Throughput must be estimated for heavy load conditions. To establish this value, estimations of mean values is not sufficient in that it provides no insight in the variance that is likely to occur during busy hour. A more advanced statistical method is required to cover excess situations that may occur with a specific probability. This was acknowledged in [1], where Monte Carlo methods were applied to forecast aggregate subscriber demands. To target specific Access Network scenarios, the model described here directly relates the number of simultaneous streams to local demographics and behavioral aspects such as sharing of devices and secondary device use. Moreover, here the impact of immersive Virtual Reality (VR) is targeted as an application that may disrupt bandwidth consumption in the coming decade.

What exactly is targeted by the Maximum Sustained Throughput is illustrated in Figure 1, where the hourly average traffic levels (grey on the left) is considered at the busy or busiest hour of the day (blue). Within that hour, we look for the minimum traffic level at which the users can make use of all network services for a specified quality of service, i.e. either with sufficient speed to ensure glitch-free video streams, or a web page refresh within a specified response time (green). Note that this Sustained Throughput level is resolved by an instant packet stream at a lower level, typically on/off switched at the peak or line rate of the access network (red on the right) or, most commonly, the subscribed rate. The Maximum Sustained Throughput is obtained from a percentile, in this study the 99th, of Sustained Throughput levels during busy hour.

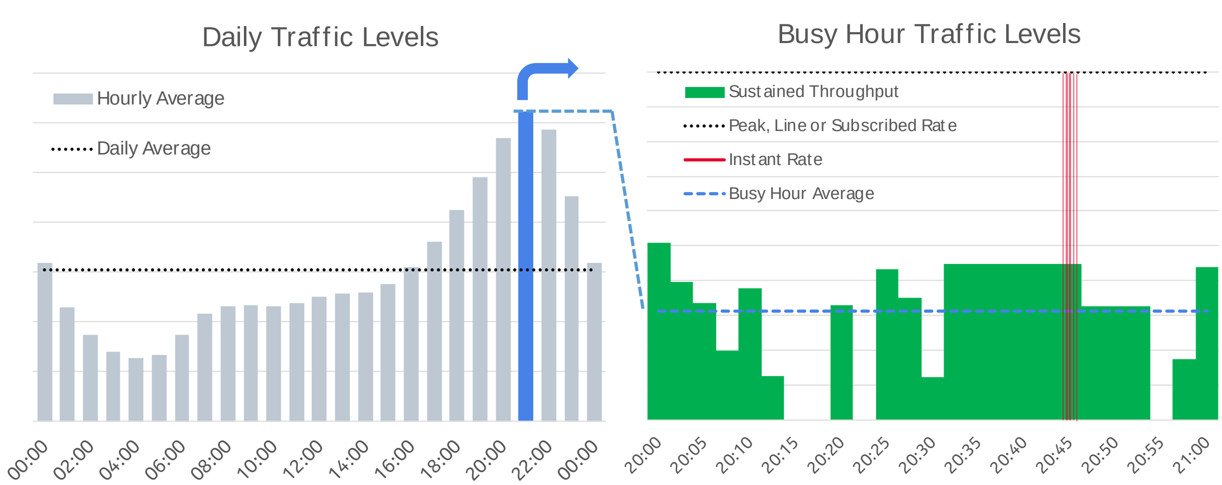

Figure 1 - Definition of the Sustained Throughput and its relation to daily average throughput (left), hourly average and instant rate (right).

Forecasting the load per home connection is difficult, as it is based on a combination of various devices and user applications that evolve constantly over time, and user generations, old and new, that change their habits and preferences. Although national or global traffic forecasts can provide a high-level glimpse of the future by extrapolating historical growth rates [2], they are insufficient for access network designs. Access networks typically depend on features of (future) technologies that, in addition, are commonly deployed selectively to meet specific local markets and geographic conditions. Network usage by users has many dependencies whose effect is hard to determine today, let alone 10 years from now. Incorporating multiple parameters into the model adds complexity. As these values are difficult to obtain, such a model may not add any more accuracy. A more deterministic model that is tractable could prove more accurate. The model for the Sustained Throughput per household B_{HH} distinguishes between two types of traffic corresponding to the way data users consume data, either through streamed media or instant downloads:

$$B_{HH} = B_S + B_D$$
$$with\ B_{S/D} = \sum_{N_{S/D}} R_{s/d} \tag{1}$$

- Traffic from media streams, B_S, is directly consumed by users and originates from multiple active traffic streams N_S at a rate R_s. Typically, this involves video and audio either on dedicated devices or as part of (web) applications or games. N_S is determined by the number of users and their usage probability which is depends on the duration of use.
- Traffic from N_D consists of instant downloads of content amounting to B_D, and is initiated either by explicit user requests ranging from chat and email to file or web page downloads, or by automated processes such as background storage backup and file sharing. Usage probability here is determined by the duration of bursts rather than the actual use duration.

The Sustained Throughput is obtained by calculating the number of concurrent streams and downloads, multiplied by their corresponding rates. The Maximum Sustained Throughput is calculated for high-load conditions, i.e. during Busy Hour. Since the number of streams and downloads depends on the usage probabilities, not only the expected mean can be quantified, but also median value and higher percentiles.

Note that most services and applications can involve a combination of both traffic types previously described. Game applications can make use of media streams, content downloads as well as periodic updates of scene data, while email has both user initiated downloads and background updates. In any case, the number of active generators N_S and N_D is determined by the number of persons per household and their use of various applications that generate them.

1.1. Population

Household size projection is based on extrapolation of U.S. Census demographic data [3], as shown in Figure 2. Here, the 2016 household size is extrapolated by 10 years using the 2014 to 2016 average growth rates for each size while applying U.S. population projections [4]. This shows a drop from 2.46 persons per household in 2017 to 2.36 in 2027. Note that household sizes may differ significantly from the U.S. average for specific cases involving other countries, states, and various residential areas.

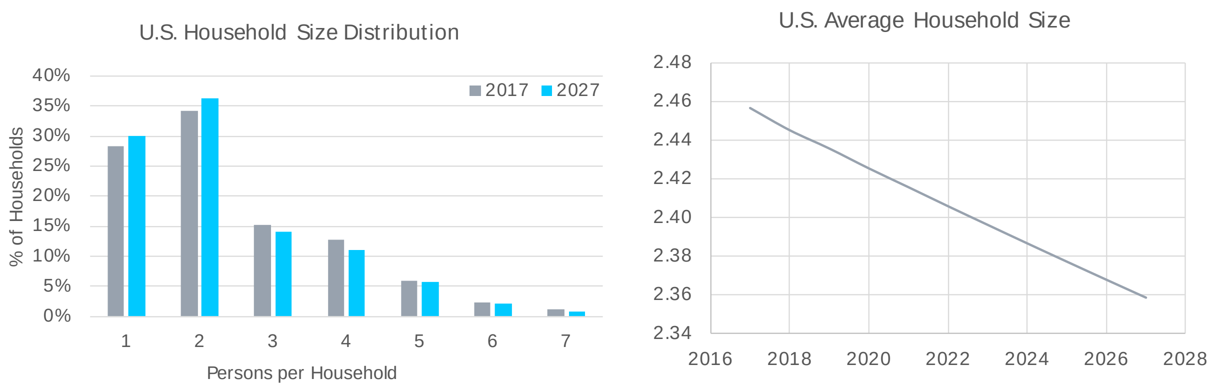

Figure 2 - U.S. Household Size Distribution (left) and Average Size (right) between 2017 and 2027, from [3,4].

1.2. Application Usage

Application usage is applied as a Usage probability p_u per resident and can generally be obtained from published reports from the number of users N_{users} of an application or service divided by the total or studied population N_{pop}, and multiplied by the minutes of use t_{use} per time frame Δt, e.g. minutes per day:

$$p_u = \frac{N_{users} t_{use}}{N_{pop} \Delta t} \tag{2}$$

For the Maximum Sustained Throughput, however, the usage probability during Busy Hour (typ. 8-9pm) $p_u(t_{BH})$ is especially of interest, and the average usage probability per day is modulated by the usage probability per hour. As reported in [5], in 2016, U.S. adults age 18+ spent 6h per day consuming media (average between September 26 and December 25 2016), of which about 5.5h is on TVs and connected devices, 21 minutes on PCs and Tablets and only 2 minutes on Mobile Phones. Neglecting concurrent use of multiple devices for the moment, modulating the average video usage probability of 22% with the hourly profile from [6], we get Figure 3, showing the Busy Hour peek at 9PM of around 52% for TVs, PCs and tablets. Here the data for adults 18+ are scaled to estimate the usage including teenagers. Although daytime use patterns may differ for that age group, it is assumed that this is less the case during the targeted busy hour. The same applies for the actual network load: the referred report does not distinguish media use at home from elsewhere and what mobile traffic is off-loaded to the home network. Given the dominance of TV streaming and the busy hour at 9PM however, it is presumed that the traffic primarily flows through the internet connection of households.

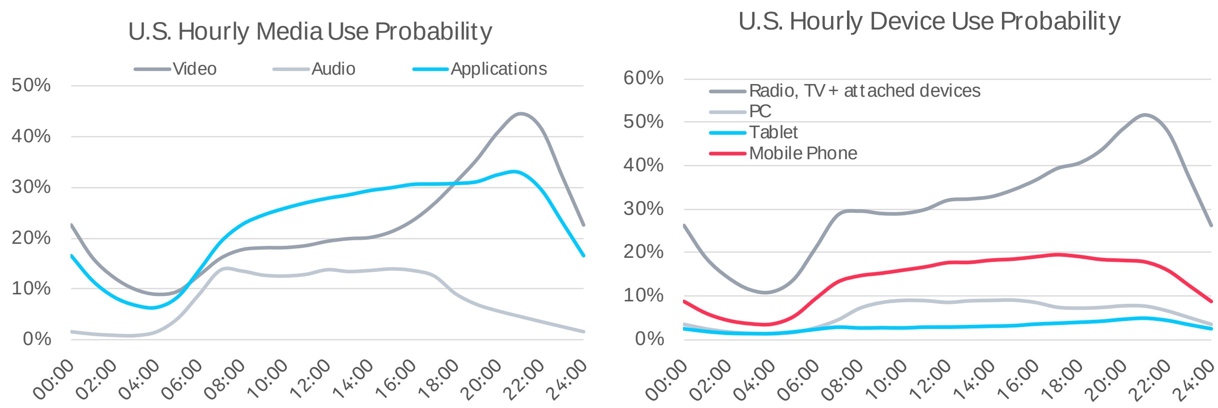

Figure 3 - Media usage (left) and device usage (right) probability per hour p_u per U.S. resident, based on [5,6].

1.3. Streamed traffic

The stream traffic per home connection is modeled as the sum of N_S active traffic streams, each contributing an amount R_s of traffic:

$$B_S = \sum_{i \leq N_S} R_{s,i} \tag{3}$$

Instead of associating traffic generators with devices or users, they are directly associated with active media streams per household. The model does require proper estimation of its two key components:

- The number of active streams N_S, and how it relates to
  - household size N_{HH},
  - the usage probability during Busy Hour p_u, but also effects such as
  - sharing of streams, such as radios and TV sets, by multiple users and
  - use of secondary streams on other devices
- The stream rates R_s, for
  - all possible device classes
  - today and in the future.

Streams can be associated with any video, audio or other media application, including gaming content and in-game communications, voice and video communications, and obviously, video. Although any media stream of any application would apply, for the maximum throughput mainly video streams (containing audio) will play a role. Typically, a distinction should be made between Access Network services that provide IPTV or other managed linear and non-linear TV (Triple Play) and those that do not. For this paper, the model assumes full TV services, either managed by the access network provider or from an Over-the-Top (OTT) subscription.

1.3.1. Stream usage

The number of concurrently active streams depends on household size and usage probability. From the data described in section 1.2 above, the video and audio usage probability can be derived as shown in Figure 4, which indicates that at prime-time almost half the population is watching video of some kind.

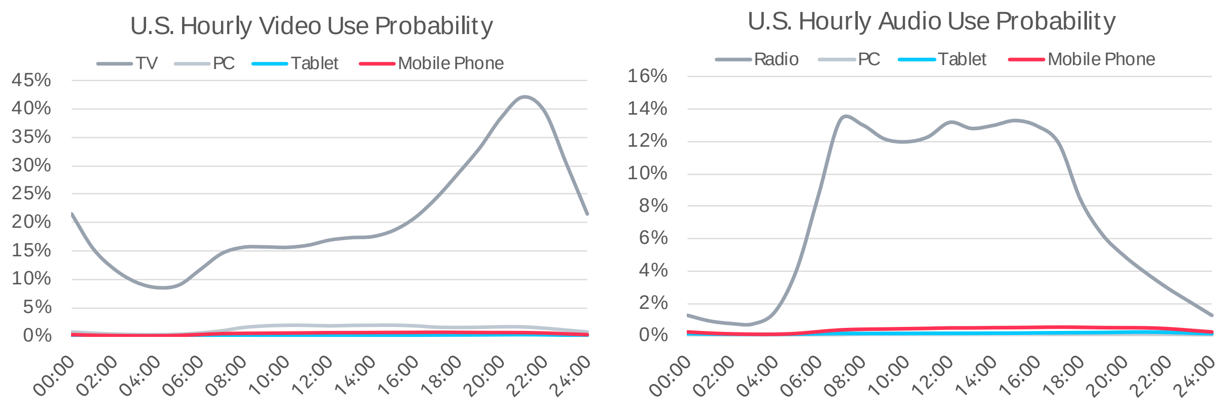

Figure 4 - Hourly video (left) and audio (right) usage probability $p_u(t)$ for U.S. residents, based on [5,6].

The number of active users N_U is modeled by applying a Binomial distribution with the household size N_{HH} and Busy Hour take rate, or usage probability $p_u = p_u(t = t_{BH})$ according to:

$$P\{N_U = k\} = \binom{N_{HH}}{k} {p_u}^k (1 - p_u)^{N_{HH}-k} \tag{4}$$

This model can be applied for individual households and, when their N_U can be assumed to be independent from each other, for Access Nodes that connect multiple households. As will be discussed later, this independence may not hold for special popular events.

To translate the number of active persons per household into the number of active streams, the number of shared views of a single stream should be observed as well as the concurrent use of multiple devices by a user (e.g. both TV and mobile phone).

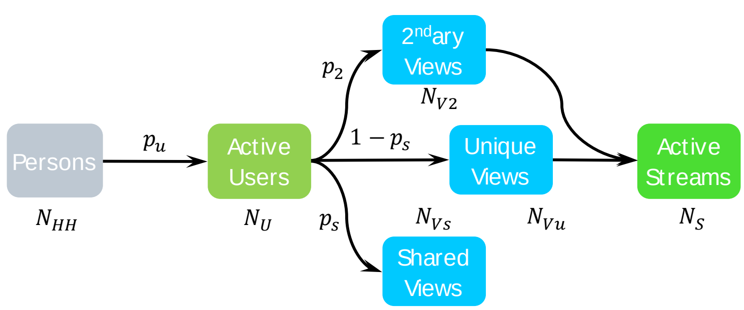

Figure 5 - Probability model to translate persons per household to active streams.

This is modelled as illustrated in Figure 5, where the number of shared streams N_{Vs} can again be modeled with sharing probability p_s for all active users that share a view. Conversely, to get the number of unique views, N_{Vu}, a Binomial distribution is used with $1 - p_s$, insofar N_U exceeds one (otherwise there is no one else to share with):

$$P\{N_{Vu} = k\} = \binom{N_U - 1}{k}(1 - p_s)^k {p_s}^{N_U - 1 - k}, \qquad N_U > 1$$
$$N_{Vu} = N_U, \qquad N_U \leq 1 \tag{5}$$

Although video is still viewed predominantly on TV sets, the number of televisions per U.S. household in fact exceeds the number of persons [7]. Therefore, sharing is assumed to involve no more than 50% of all views. By applying the probability that an active user uses a secondary stream, p_2, the number of secondary views N_{V2} can be obtained:

$$P\{N_{V2} = k\} = \binom{N_U}{k} {p_2}^k (1 - p_2)^{N_U - k} \tag{6}$$

It is assumed that this probability is low, i.e. 5-10%, and that concurrent use of more than 2 streams per user is nihil. The total number of active streams is now simply the sum of unique and secondary views:

$$N_S = N_{Vu} + N_{V2} \tag{7}$$

Unlike usage probability, the estimates for both secondary usage, p_2, and sharing, p_s, are rather speculative for now and needs proper evidence to determine to what extent they depend on user behavior, especially that of teenagers. Additionally, projections will depend strongly on new types of devices and future applications.

As indicated by Figure 6, although the average household size is expected to decrease from 2.46 persons today to 2.36 in 2027, a decline in shared usage will keep the number of active streams per household close to 1.3 between 2017 and 2027.

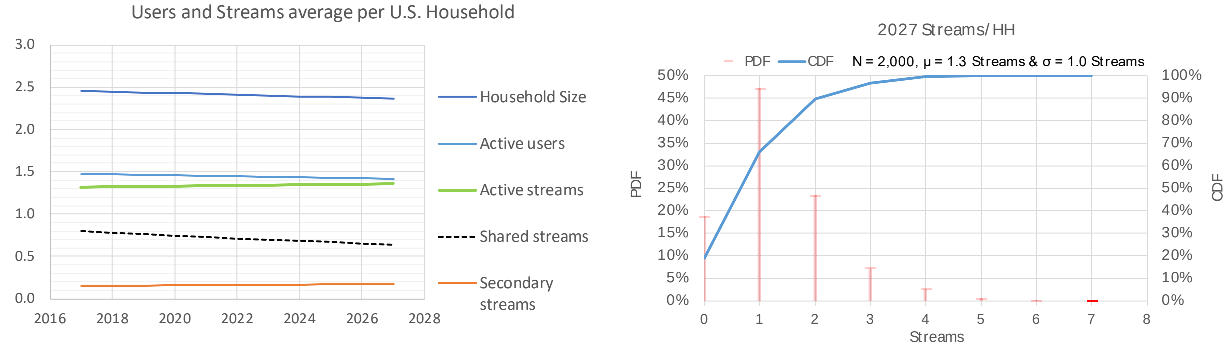

Figure 6 - Projection of persons on streams (mean values, left), accumulating into the active streams (distibution, right) per average U.S. household.

1.3.2. Stream rates

More difficult to predict is the actual traffic per stream. Shared devices, televisions and beamers, typically consume higher rates than mobile devices. Managed linear and non-linear video services often exhibit higher video rates than OTT video services, where adaptive stream rates are commonly used to limit data consumption and server loads. Tablets and phones on the other hand show higher replacement rates than TV sets, while it is hard to predict the future adoption of virtual reality (VR) devices [8]. For this reason, again a probabilistic approach is chosen to estimate stream rates based on common resolutions and their adoption in the coming years as indicated in Figure 7. For 2017, this distribution is based on 37% of Standard Definition (SD) streams below 1k, and 47% around 2k and 16% 4k [9]. The model does not make a distinction between TV sets and mobile devices, as their resolution is presumed to evolve in parallel. 8k Video is assumed to be adopted as of 2020, while 16k starts in 2025. Note that these higher resolutions would represent Ultra High Definition (UHD) streams to both big-screen TVs, projectors and, for non-interactive streams, to VR goggles. Interactive immersive VR streams are addressed separately in section 1.3.3 below.

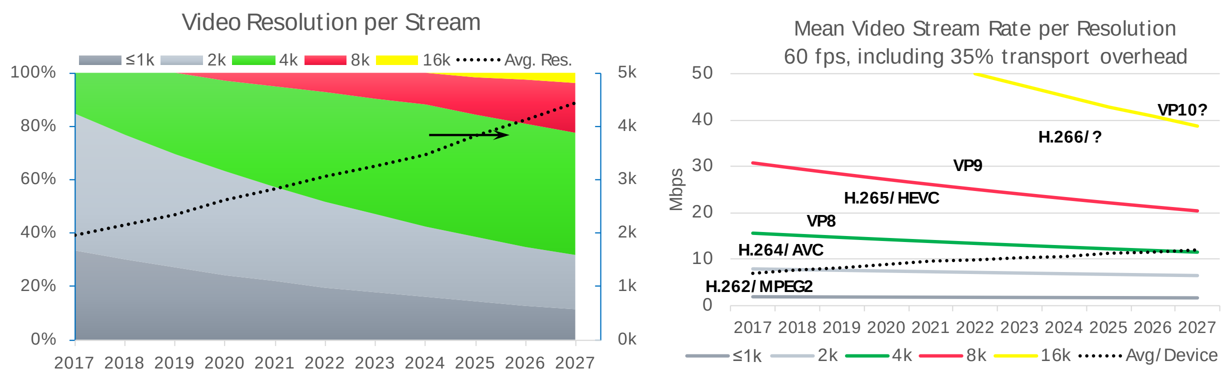

Figure 7 - Stream resolution distributions combined for current and emerging TV and mobile devices per household (left) and associated mean video rates and coding (right).

Sustained video rates for various stream resolutions will still be suppressed by using lossy compressions and codecs which in turn will evolve over time. The projections shown here are based on the analysis from [10], using the AVC and HEVC rates for 2k HD and 4k UHD video for a (subjective) quality score of 8 out of 10, and adding a transport overhead of 35%. This will apply to most managed TV and VoD services, albeit higher than many OTT streams. To account for quality and (adaptive) rate variations, the values shown in Figure 7 are varied by using a Gamma Distribution with σ/μ of 10%. Figure 8 shows an example for the 4k video sustained rate distribution in 2017. For the forecast, a compression improvement of 10% (MPEG2) up to 50% (H.266) per decade, or 1 to 5%/year is assumed, reflecting the migration to higher compression standards per year, due to device replacements and upgrades. 8k and 16k rate estimations are mere extrapolations of the lower rates, presuming that devices will have access to the processing power needed for future, i.e. H.266 or VP10, compression standards and codecs. The resulting average stream rate per household increases from 7 Mbps in 2017 to 12 Mbps in 2027 for an average resolution growing from 2k to 4.5k.

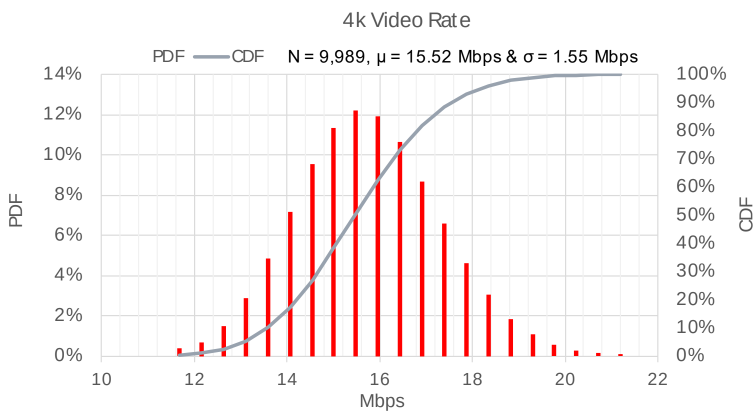

Figure 8 - 4k video stream rate variation in 2017 arround the mean value to account for quality variation and rate adaptations.

1.3.3. Interactive immersive streams

Although the projected stream rates would include 16k video resolutions that can be expected from VR headsets, one aspect that is not covered is interactivity. To enjoy immersive experiences without nausea and motion sickness, VR stream rates will likely need to support 100 frames per second and millisecond response times in rendering new frames corresponding to a movement. Although technologies are yet to be developed for real-time network streaming that allows massive adoption, a scenario with interactive VR is included to assess the throughput sensitivity to new disruptive applications.

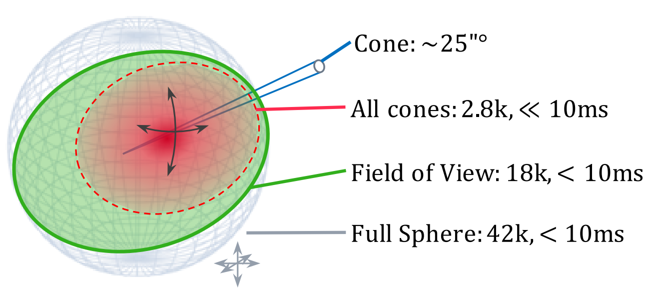

Figure 9 - Visual scope, 16:9 equivalent resolution and corresponding latency requirements for immersive VR.

To estimate potential video rates for immersive virtual reality applications, the basic properties of human vision are illustrated in Figure 9. Presuming perfect eye quality, the cones in the central foveal area are taken as a reference [11]. Translating 4.5 million cones per eye to a standard 16:9 ratio screen resolution equivalent $(4/3\sqrt{N_{pixel}})$ yields a modest 2.8k video stream. For a stereoscopic image, assuming 50% overlap and compression efficiency, this would result in a 4k stream. However, expanding a perfect retinal acuity of 20 to 30 arc seconds per cone to the full field of view (FoV, ca. 145° wide x 100° high) yields an equivalent resolution of 18k per eye. This means that to anticipate any possible eye movement you need to transfer 8 times as much information as a user actually can perceive within a FoV (similar for screen displays). By using eye tracking, the adaptive foveal stream of 4k would suffice, but then response times far below 10ms are needed, including latency of viewing gear, server and transmission. Since directional head and body movements are slower than eye movement, response times can be more forgiving but still within 10 milliseconds to prevent motion sickness. A full spherical view ($360°^2/\pi$ or about one billion hexagonally arranged cones) can instantly accommodate any change of view direction, but then an equivalent stereoscopic resolution of about 60k is needed. Obviously, when the viewer also moves around, physically or virtually, latency requirements will also apply here. When full sphere image streams can be compressed as sufficiently as HEVC projections promise then, for frame rates in the order of 100 fps, data rates around 200 Mbps can be expected. Since compression of spherical images will ultimately depend strongly on scenery, compression algorithms and certainly on device processing capabilities, both 200 and 400 Mbps is investigated. A second important parameter is the expected use probability. Although daily use patterns will probably resemble other video applications, the adoption of new devices is very uncertain. TV headset penetration rates up to 45% are expected [12] in the coming 5 years, but this includes smartphone-based devices. For stand-alone devices with capabilities projected here (18k per eye), scenarios are included for adoption as of 2020, growing to either 7% or 15% in 2027.

1.4. Download traffic

To translate the number of persons per household into the number of active downloads, the concurrent multiple use of devices (e.g. TV, PC, tablet and mobile phone) should be observed as well. For example, a member of the household may be using their PC to check email, browse the web or go on social media and then use their tablet or smart phone to do the same activity within the same busy hour time interval. The traffic generated in the household by instant content downloads can similarly be modeled as a sum of generators:

$$B_D = \sum_{i \le N_D} R_{d,i} \tag{8}$$

The number of concurrent downloads N_D can be expressed again as:

$$P\{N_D = k\} = \binom{N_{HH}}{k} {p_u}^k (1 - p_u)^{N_{HH}-k}$$

According to (2), the download probability p_u is determined by the use duration. But unlike streaming content, this is the download or burst duration rather than the application use duration which, for a daily download size D per user, amounts to:

$$t_{use} = \frac{D}{R_d} \tag{9}$$

User initiated downloads do however depend on device use, so the hourly download probability of applications $p_u(t)$ can be derived by distributing the data consumption per device by the device use per hour. From the data described in section 1.2 above, Figure 10 results for various devices. For non-interactive data applications that do not rely on a specific response time, i.e. from autonomous devices and processes such as storage and backup, a constant uniform distribution profile is used.

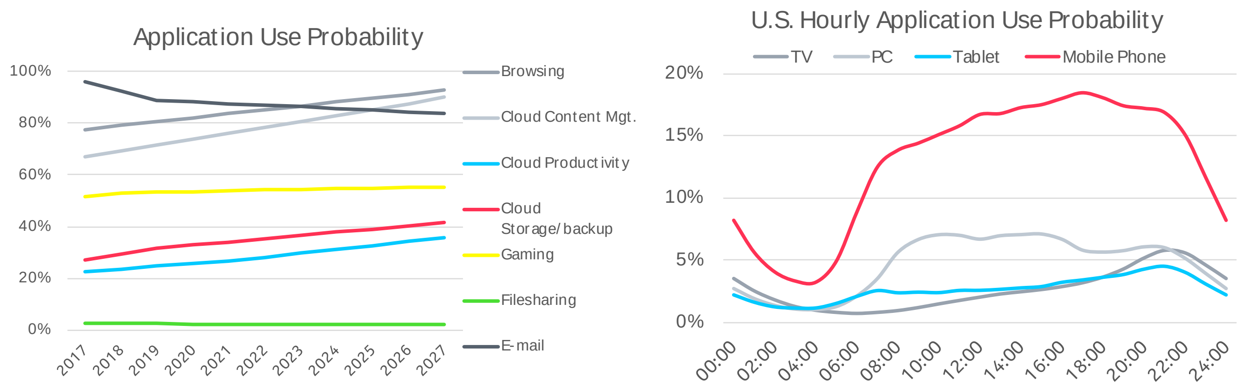

Figure 10 - Application usage probability p_u per U.S. resident (left) and hourly distribution per device, based on [5,6] (right).

1.4.1. Download Rate

The burst rate R_d in (8) of application downloads is, apart from the response time t_r given by the burst data size d_b:

$$R_d = \frac{d_b}{t_r} \leq R_{max} \tag{10}$$

For unconstrained use one could argue this rate should be as high as possible, i.e. the maximum rate supported by both home network and access network R_{max}. But more realistically, also at the server side restrictions will apply, so that two principles should be observed. First, the amount of data that users can consume instantly is limited by perception restrictions. Most often, visual information in the form of shapes, images and photographs will be the most demanding (non-streamed) content, outnumbering plain text data by a factor of 1,000 to 10,000. Secondly, interactive view and browser applications apply progressive download techniques to present content before full download has completed. This applies to images, web pages and documents but also to interactive table views to online data sources.

In this study, the burst size is assumed to amount to 2.5 MB for an average web page [13], increasing 6% per year, parallel to the average stream rate (see Figure 7) while the response time is assumed 4 seconds (target web page download time [14]) decreasing to 0.5 seconds in 2027. Other application burst sizes and durations are merely estimates for typical applications and content types. But similar to stream rates, burst sizes will vary strongly and are therefore modeled using a Gamma distribution with a σ/μ of 50%.

With the burst rate, also the download probability can be established from (2), (9) and (10):

$$p_u = \frac{N_{users}}{N_{pop}} \frac{D}{\Delta t R_d} = \frac{N_{users}}{N_{pop}} \frac{D}{\Delta t} \frac{t_r}{d_b} \tag{11}$$

With $D/\Delta t$ the daily download size per user. As indicated, the burst probability is the product of user probability N_{users}/N_{pop}, the number of burst per day D/d_b and the time fraction of a burst $t_r/\Delta t$.

2. Calculations

2.1. Household traffic model

Although the model described allows for direct derivation of mean throughput values, the aim here is to assess the variation resulting from all assumed distributions. Especially for smaller Access Distribution Areas (DAs), e.g. for FTTdp and FWA, the network is commonly designed for a certain percentile well above the expected average, e.g. the 95th percentile. The cascade of probability distributions for users, usage and video rates prevents expressing the n-th percentile of the throughput directly. Although it is possible to scan through all permutations numerically to collect percentile values, here we revert to Monte Carlo simulation. For this purpose, thousands of households are calculated (N = 10,000 for each year) by using random samples for the various distributions. To suppress numeric noise when calculating low-probability conditions of high-percentiles, Latin hypercube (LHC) pseudo-random sampling is used to cover most of the variation ranges of population, device and stream distributions. A summary of the main model parameters is listed in Table 1 and Table 2.

Table 1 - Media stream model parameters

Parameter	**Value**
Busy Hour Usage probability p_u	μ = 45%, σ = 23%
Secondary Streams per Active User p_2	10% ± 10%, CAGR +2%/year
Shared Stream fraction p_s	50% ± 50%, CAGR -2%/year
Video rates per stream (@60 fps, 35% overhead) R_i:	Gamma distributed, σ/μ = 10%
• ≤1k	μ = 2 Mbps, CAGR -1%/year
• 2k	μ = 8 Mbps, CAGR -2%/year
• 4k	μ = 16 Mbps, CAGR -3%/year
• 8k (as of 2020)	μ = 27 Mbps, CAGR -4%/year
• 16k (as of 2025)	μ = 43 Mbps, CAGR -5%/year
• Interactive VR (if included, as of 2025)	μ = 200/400 Mbps
Audio BH Usage probability p_u (Radio & VoIP)	μ = 8.8%, σ = 4.4%
Audio rate per stream (Radio & VoIP)	0.128 Mbps

Table 2 - Data download model parameters

Application	Users per resident	Data Use D MB/day	Burst Size d_b [MB]	Burst Time t_r [s]
Browsing	76%, CAGR +2%/y	89.1 +8%/y	2.5 +6%/y	4 -21%/y
Content	65%, CAGR +3%/y	46.5 +15%/y	5 +6%/y	4 -21%/y
Productivity	22%, CAGR +5%/y	97.2 +14%/y	10 +6%/y	4 -21%/y
Storage/backup	24%, CAGR +5%/y	84.7 +8%/y	50 +6%/y	300 -21%/y
Gaming	50%, CAGR +1%/y	10.6 +5%/y	0.01 +6%/y	0.1 -21%/y
File sharing	3%, CAGR -3%/y	2,700 +2%/y	200 +6%/y	600 -21%/y
E-mail	99%, CAGR -2%/y	3.1 +3%/y	5 +6%/y	10 -21%/y

2.2. Distribution Area traffic model

For Distribution Areas, the aggregate throughput of a collection of connected households is calculated by resampling the data obtained per household. This is done to limit computation time without having to rely on less accurate fitting to, e.g., Gamma distributions.

2.2.1. Multicasting

The total aggregate throughput at the line side of the Access Node is calculated as the sum of all connected home throughputs, where it is assumed that no multicasting mechanisms are available at the link layer. For throughput at the uplink of an Access Node, but also access technologies on shared media that do support link multicasting, e.g., HFC, PON and FWA, throughput calculations of linear TV streams must account for savings from broadcast streams. The same would also apply for IP multicast on Access Node uplinks, depending on group numbers and stream popularity distributions. Increased use of Unicast streams, i.e. for non-linear and time-shifted viewing, will reduce multicast savings. When Access Nodes provide caching however, uplink transport savings can be obtained as indicated by the Hit Ratio. Figure 11 shows content popularity, commonly modeled as Zipf-Mandelbrot distributions and the resulting Multicast Gain that can be obtained. For typical multicast and VoD shape parameters, 80% of multicast traffic can be saved for 1000 views assuming 250 channels. For VoD and time-shifting, usually more content is available so that, for 2000 titles, 30% of unicast traffic could be saved for 1000 viewers. If unmanaged OTT streams are cached at all, this mostly takes place deeper in the network, without impact on the load of Access Nodes or their uplinks. Although Figure 11 suggests that significant load reductions can be achieved especially for multicast video in larger Distribution Areas, this effect has not been included in the remainder of this analysis.

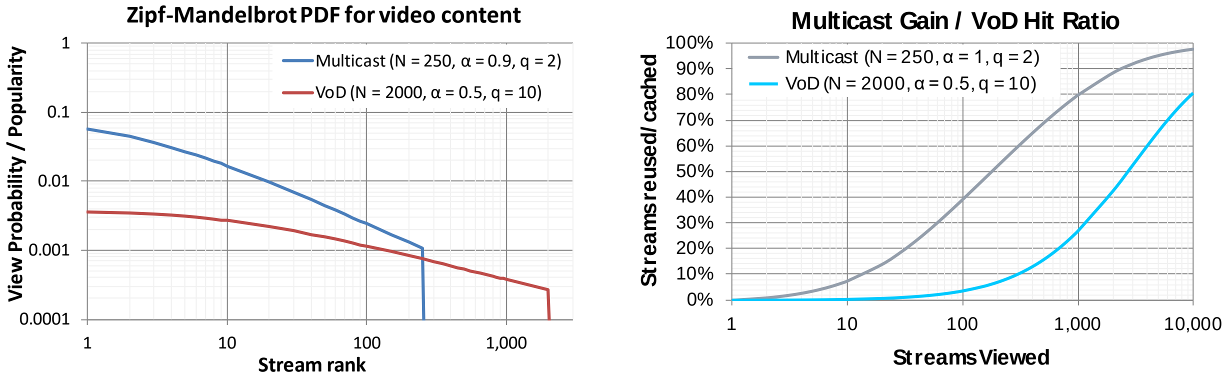

Figure 11 - Managed content popularity distributions (left) and corresponding Multicast Gain and Cache Hit Ratio (right).

2.2.2. Usage dependence

Although resampling provides an efficient way of calculating aggregate throughputs based on the statistics per household, it disregards one aspect that may play an important role. It implicitly assumes independence of the usage probability among different households which, especially for TV video streams may not apply in case of popular sports or newsworthy events. To see the impact of such events, an alternative scenario with a view probability of 90% instead of 45% is calculated which corresponds to the total population of 10 years and older watching TV at 9pm.

3. Results

3.1. Sustained Throughput per U.S. household

3.1.1. Mean Sustained Throughput

Using the distributions and assumptions discussed in sections 1.3, 1.4 and 2, the Sustained Throughput is calculated for each year. Figure 12 shows contributions of streams and downloads to the mean throughput values during Busy Hour between 2017 and 2027, which are summarized in Table 3. The slight curving at 2020 and 2025 correspond to the uptake of 8k and 16k video respectively. The average throughput in 2017 is about half the BH rate and would amount to 4 Mbps. As a reality check, this corresponds to 1.3 TByte of video data per month, increasing to 2.4 TByte/month in 2027. Again, this would represent a situation where linear TV services are taken either from the Access Network provider or from another IPTV or OTT TV subscription.

Table 3 - Mean Sustained Throughput per U.S. Household

Mean Sustained Throughput per HH	2017	2027	CAGR
BH Media Streams	7.9 Mbps	14.2 Mbps	6.7%
BH Data Download	0.2 Mbps	0.7 Mbps	12.9%
BH Total Throughput	8.1 Mbps	14.8 Mbps	6.9%
Daily Average Throughput	4.0 Mbps	7.4 Mbps	7.0%
Data per day	43.2 GB	79.5 GB	7.0%
Data per month	1.3 TB	2.4 TB	7.0%

These numbers indicate almost a doubling of growth from 8 to 15 Mbps per household. This results from the increased number of streams per household (Figure 6) and especially the mean rate growth (Figure 7). The impact of streaming video is clear, as it contributes to 95% of the total volume per household.

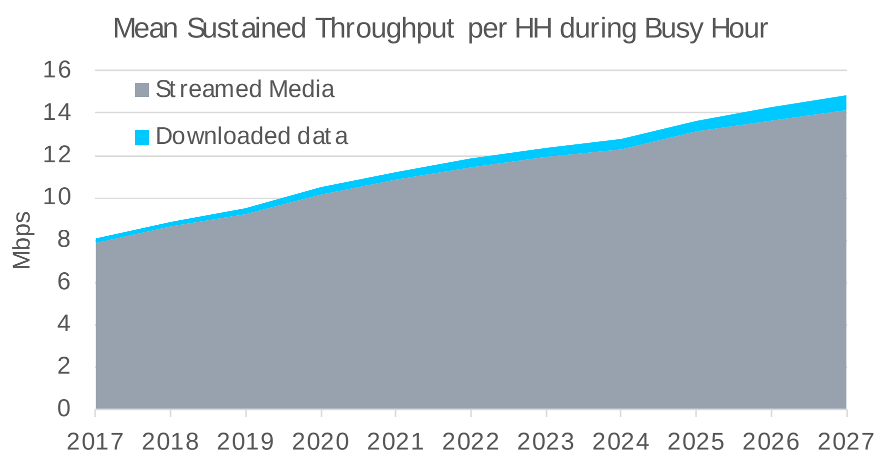

Figure 12 - Mean Sustained Throughput per U.S. average household contributions from Streamed Media and Data downloads during busy hour.

3.1.2. Throughput Distributions

Using the distributions and assumptions discussed in sections 1.3, 1.4 and 2, the Sustained Throughput is calculated for each year. Figure 13 shows the results for 2017 and 2027, indicating a growth of the mean Throughput value from 8.8 to 15.6 Mbps per household. This results from the increased number of streams per household (Figure 6) and especially the mean stream rate growth (Figure 7). The Probability Density Functions (PDF) show the long tail shape as they propagate from the Exponential-like shapes of both the stream number and size distributions.

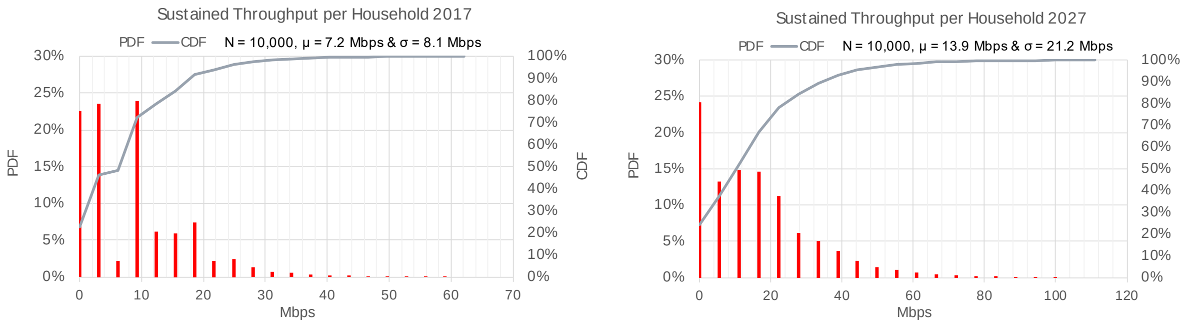

Figure 13 - Probability Density Function (PDF) and Cumulative Density Function (CDF) for the Maximum Sustained Throughput per U.S. household in 2017 (left) and 2027 (right).

3.1.3. Percentiles vs Mean value

A more significant value for network design is the upper range of the expected throughput values. Figure 14 and the summary in Table 4 show the various percentiles for the Throughput values between 2017 and 2027, indicating that to design for e.g. 95% of all households in 2027, a throughput of up to 43 Mbps per household would need to be considered rather than the mean value of 14.8 Mbps. For 99% this value reached 71 Mbps. For an Access Node, the difference between mean and high variations depends on how many households are actually connected per Access Node, and thus the Distribution Area size as analyzed next.

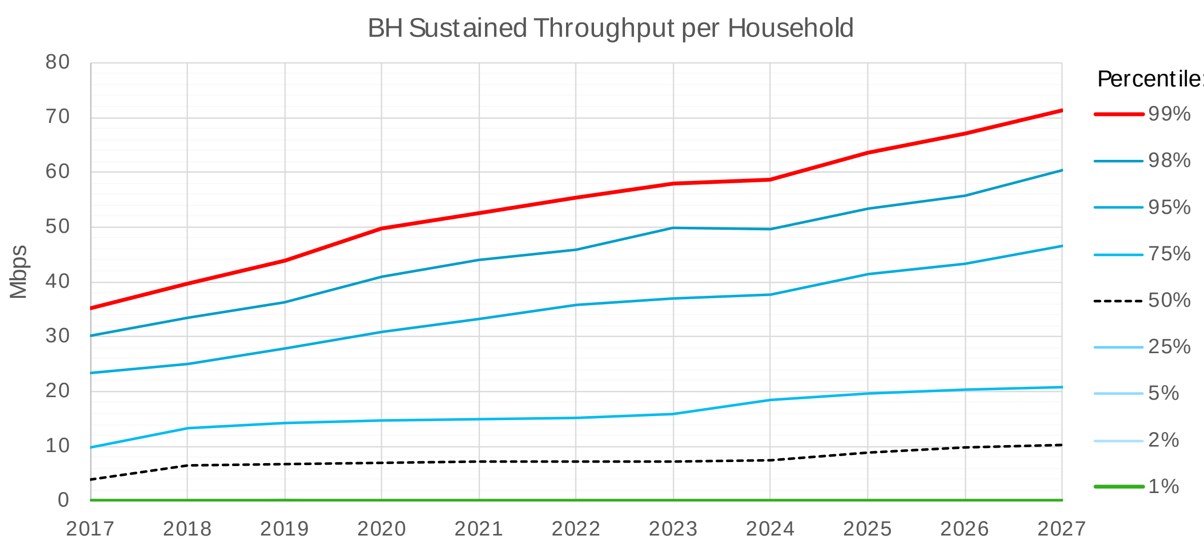

Figure 14 - Percentile range for the Sustained Throughput per U.S. household between 2017 and 2027.

Table 4 - Sustained Throughput percentiles per U.S. Household

Sustained Throughput per HH	2017	2027	CAGR
Mean	8.1 Mbps	14.8 Mbps	6.9%
50% (Median)	2.5 Mbps	10.3 Mbps	17.3%
95%	21.8 Mbps	46.5 Mbps	10.0%
99%	35.1 Mbps	71.4 Mbps	11.6%

It may be noted that networks in areas with higher revenue potentials will have to be designed to cover 95th or even 99th percentile throughputs as residents will expect service quality comparable to FTTH. Additionally, sufficient headroom is to be kept at distribution points as margin on top the sum of sustained throughput of all users, in order to cater for bursty applications or “speed tests” that users may perform from time to time to check their experience versus advertised headline speeds.

3.1.4. Stream use Sensitivities

To assess the model sensitivities to the major parameters, results from the simulations are shown in the scatter plots of Figure 15 and Figure 16. As expected, the Busy Hour usage (Figure 15, left) linearly drives the number of streams and thus the throughput per household. A more detailed analysis of high traffic loads during special occasions is analyzed in section 3.2.1 below.

The use of secondary simultaneous streams (Figure 15, right) has a much lower impact as its range is a factor 4 smaller than that of p_u and it impacts the number of active streams only relatively.

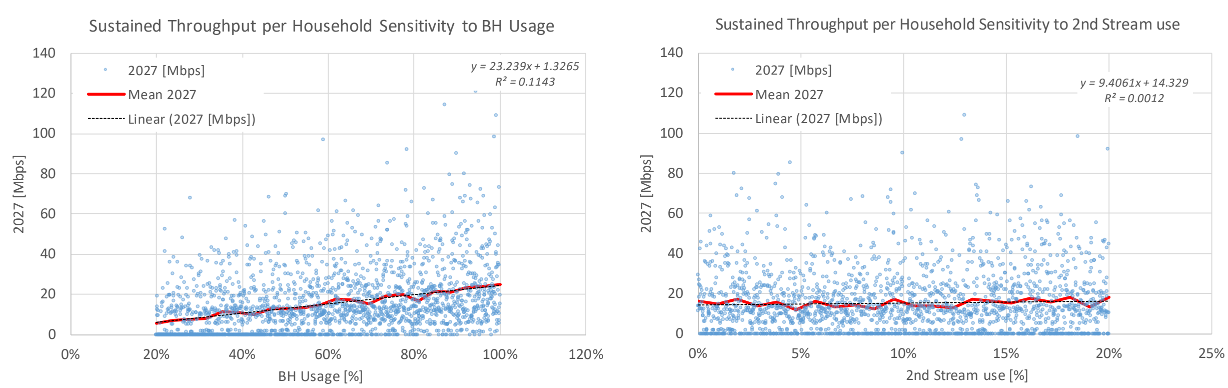

Figure 15 - Sensitivities of the Sustained Throughput per U.S. household in 2027 to variations in Busy Hour Usage p_u (left) and secondary stream use p_2 (right).

As indicated by Figure 16, the effect of Stream Sharing is more visible, but smaller than (BH) Usage: it only dampens active streams when more than one active user is present. Since an average household has only 1.3 active viewers during busy hour (45% of 2.45 persons), sharing only applies to 0.3 views per household or 25%.

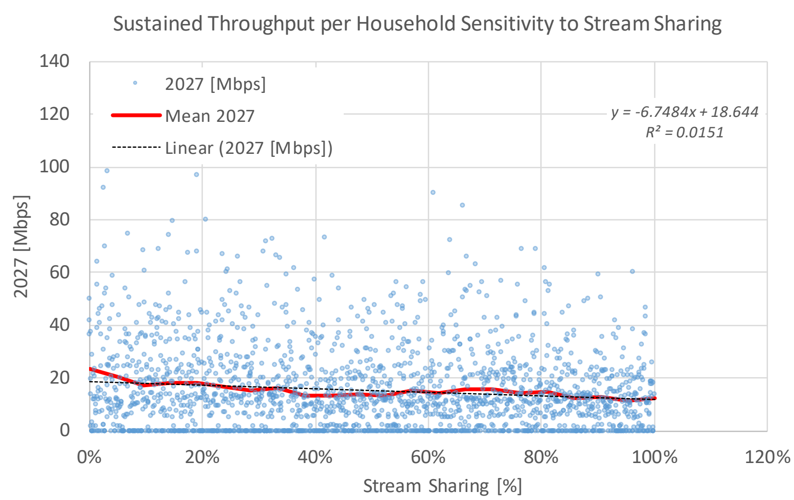

Figure 16 - Sensitivity of the Sustained Throughput per U.S. household in 2027 to variations in Stream Sharing p_s.

3.1.5. Disruptive applications

To determine the impact of new disruptive applications, a scenario has been calculated where 200 Mbps streams associated with immersive, interactive VR were added to the portfolio increasing from 0.1% of the video streams in 2020 to 7% in 2027. The results are shown in Figure 17 and summarized in Table 5. The Mean Sustained Throughput doubles to 15.6 Mbps, the 99th percentile Maximum Sustained Throughput jumps from 70 to 241 Mbps. This suggests that even for a low uptake of applications such as Immersive VR, the Maximum Sustained Throughput will be impacted. Higher take rates of these services can dramatically increase throughput requirements.

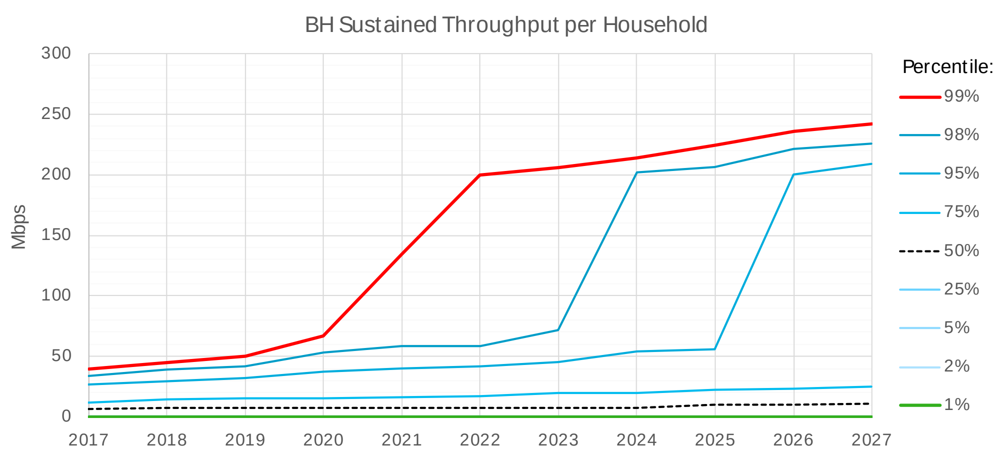

Figure 17 - Sustained Throughput per U.S. household during busy hour including interactive VR Streams for take rate up to 7% in 2027 and 200 Mbps per VR stream.

Table 5 - Impact of Immersive VR streams on Sustained Throughput per U.S. Household (200 Mbps VR rate)

2027 Sustained Throughput per HH	without VR	with 7% VR	15% VR
Mean Average Throughput	7.4 Mbps	15.6 Mbps	23.7 Mbps
Mean BH Throughput	14.8 Mbps	31.9 Mbps	48.3 Mbps
95%	46.1 Mbps	207 Mbps	223 Mbps
99%	70.4 Mbps	242 Mbps	419 Mbps

Figure 18 shows the situation when take rates are doubled to 15% in 2027. Although a higher take rate increases the throughput probabilities, it is not until 2027 when the probability of two simultaneous VR streams pushes the 99th percentile Maximum Sustained Throughput beyond 400 Mbps.

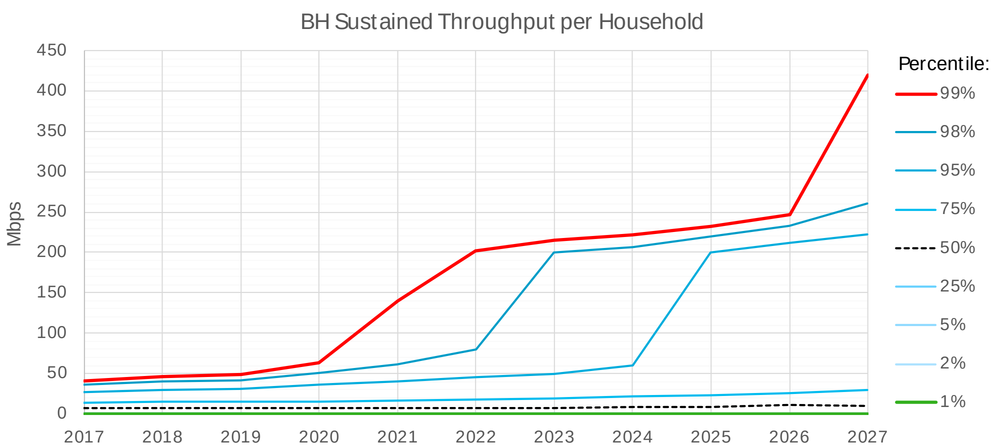

Figure 18 - Sustained Throughput per U.S. household during busy hour including interactive VR Streams for a 15% take rate at 200 Mbps per VR stream.

That these maximum rates are directly driven by the VR rate is shown in Figure 19, where a 400 Mbps VR rate is assumed, which basically doubles the Maximum Sustained Throughput. As such, close monitoring of how VR gear capabilities will translate into stream rates is essential in the coming years.

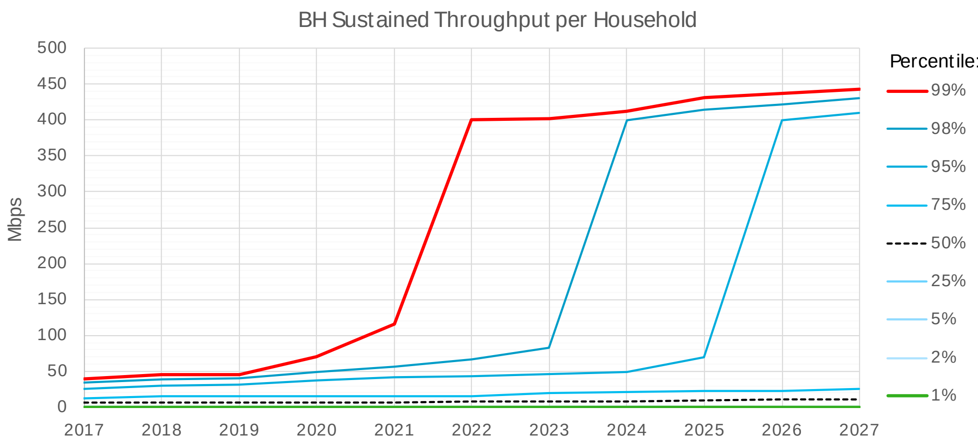

Figure 19 - Sustained Throughput per U.S. household during busy hour including interactive VR Streams for a 7% take rate at 400 Mbps per VR stream.

3.2. Sustained Throughput for Distribution Areas

Applying the per household throughput distribution to simulation of a collection of connected households gives the results shown in Figure 20. As the size of DAs increase, averaging between small and large households and various devices results in less deviation and, as the central limit theorem (CLT) predicts, a throughput distribution that resembles a Normal distribution.

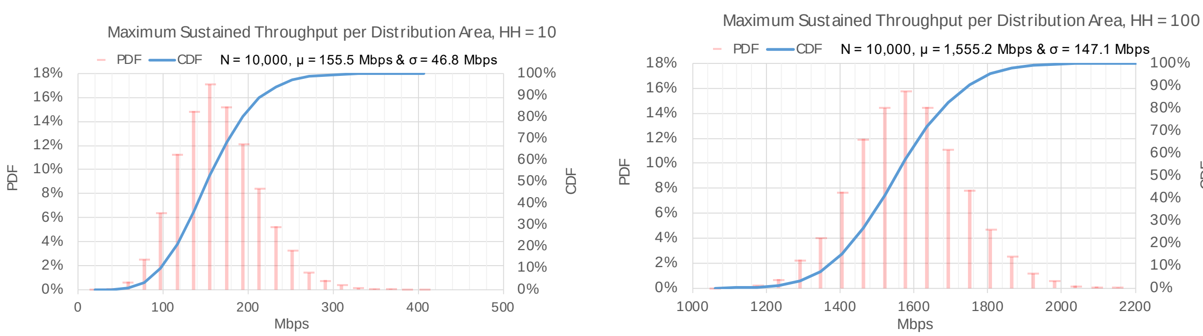

Figure 20 - 2027 Throughput for a Distribution Area with 10 (left) and 100 households (right).

This presumes some independence of the connected households which in practice, for specific geographic areas characterized by homogeneous income patterns, may actually-not be entirely satisfied and should be accounted for by selecting the proper household size distribution and possibly adjusting the stream distribution. For this study, relying on U.S. averages, the impact of DA size on the total Sustained Throughput is depicted in Figure 21. It shows that for smaller areas, i.e. < 100 households, the relative variance σ/μ is too high to ignore.

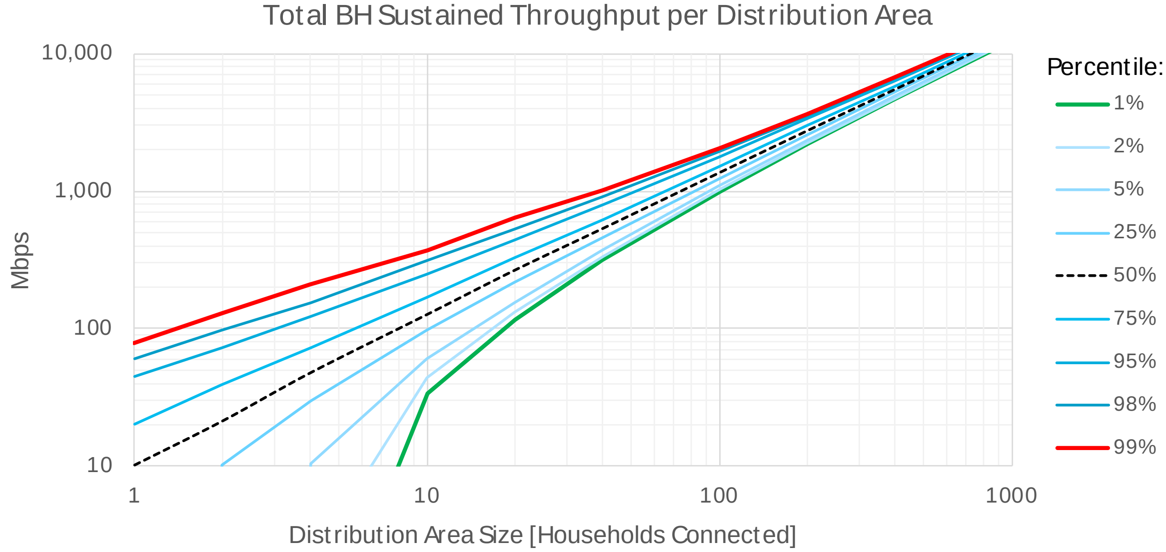

Figure 21 - 2027 Throughput for varying Distribution Area size per percentile

For N=10, the 95th percentile value is 80% higher than the mean value where for N=100, this is only 23%. For $N \gg 100$, the variance drops further and designs can assume mean throughput values similar to core transport networks. This is illustrated in Figure 22, where the DA capacity per connected household drops from 70 Mbps, the Maximum Sustained Throughput, for 1 household to 15 Mbps, almost the Mean Sustained Throughput per household.

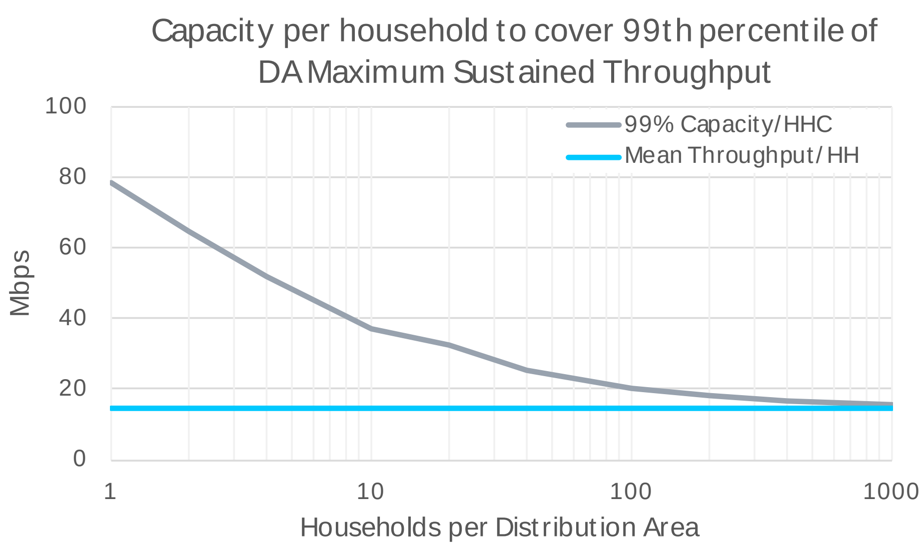

Figure 22 - 2027 Capacity per household required to support the 99th percentile of the total Sustained Throughput in Distribution Areas

3.2.1. High-load conditions

To see the impact of massive views during special sports or other newsworthy events, a scenario is calculated with a double view probability of 90%, i.e. 9 out of 10 residents watching a video stream concurrently. As indicated in Table 6, a high view probability increases the mean throughput by 75% both for the daily average and Busy Hour. The Maximum Sustainable Throughput increases only by 27%, because these include households with more viewers, and the sharing of streams exceeding 1 is more likely in big households than in average sized households. Since it is unlikely that 90% of the population is watching video concurrently even during rare events, an additional 30% can be regarded as a solid margin to design for these circumstances.

Table 6 - 2027 Sustained Throughput per U.S. Household during high load events

2027 Sustained Throughput per HH	45% Viewers	90% Viewers	increase
Mean Average Throughput	7.3 Mbps	12.7 Mbps	74%
Mean BH Throughput	14.8 Mbps	25.9 Mbps	75%
95%	46.5 Mbps	60.5 Mbps	30%
99%	71.4 Mbps	90.6 Mbps	27%

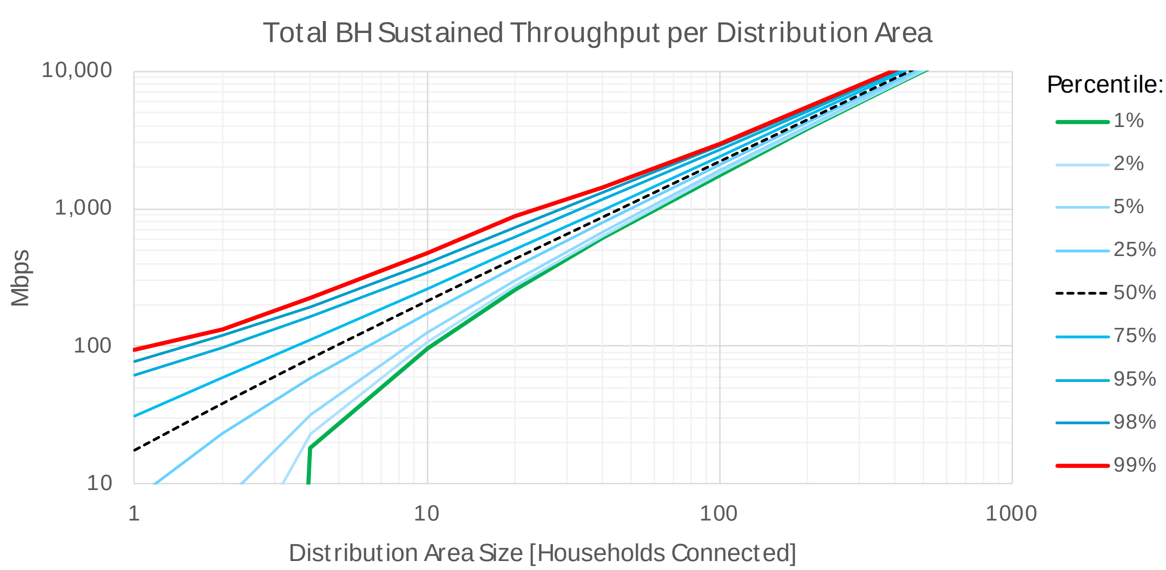

Figure 23 - 2027 BH Throughput variation per Distribution Area size during high load events.

Conclusion

Video will continue to drive residential access network loads in the coming 10 years. While peak throughputs per household may cross 1 Gbps, the mean expected sustained capacity per household by 2027 will not likely exceed 7.5 Mbps on average and 14 Mbps during Busy Hour, but that will not support the Maximum Sustained Throughput of households.

The 99th percentile Maximum Sustained Throughput per household will be 70 Mbps in 2027 during Busy Hour but during high-load conditions, when most of the population is watching video concurrently, this will increase up to around 90 Mbps. The margin to consider for network design is therefore 30% higher than normal viewing conditions.

When disruptive interactive applications such as Immersive VR will represent only 7% of video streams, the Maximum Sustained Throughput may increase to 240 Mbps or, for 15% take rates, to 420 Mbps in 2027. Most of this rate will depend on new interactive and immersive video services and eventual device capabilities. Higher adoption of high end applications can dramatically increase the service requirements.

Table 7 - 2027 Sustained Throughput per U.S. Household for engineering considerations

Maximum Sustained Throughput (99th percentile Busy Hour)	2017	2027
U.S. household	35 Mbps	71 Mbps
U.S. household - High Load	45 Mbps	91 Mbps
U.S. household – 7% Disruptive VR	-	242 Mbps
U.S. household – 15% Disruptive VR	-	419 Mbps

The throughput variance per Access Node rapidly drops with higher number of homes connected, allowing taking advantage of statistical multiplexing. The standard deviation of the Throughput for 100 connections is only a third compared to 10 connections, while 1000-household DAs can design for only a few percent above mean average Busy Hour values. However, it is important to engineer Access Nodes with sufficient headroom to accommodate at least one “Speed Test” on top of sustained throughput, in order to meet customer expectations as well as any regulatory requirements in the foreseeable future. Many regulatory authorities in Europe have discussions on “truth in advertisement” in the public domain.

Scale is therefore key as fiber pushes further towards the home and Distribution Areas shrink with growing speed. For Access Networks, either HFC, FWA, FTTH or xDSL, proper traffic forecasting is the starting step for a proper design. With the appropriate demographics indicators and (video) service projections, the presented model is well suited to provide key insights into the Maximum Sustained Throughput and other traffic characteristics. As the projections for future video usage, application bandwidth requirements as well compression techniques emerge, the current projections will evolve over time.

Abbreviations

Acronym	Meaning
5G	5th Generation (Wireless Systems)
AR	Augmented Reality
AVC	Advanced Video Coding
BH	Busy-Hour
CAGR	Compound Annual Growth Rate
CDF	Cumulative Distribution Function
CPE	Customer Premises Equipment
DA	(Access) Distribution Area
FoV	Field of View
FTTdp	Fiber to the Distribution Point
FWA	Fixed Wireless Access
HD	High Definition
HEVC	High Efficiency Video Coding
LHC	Latin hypercube
Mbps	Megabits per second
μ	Mean Average Value
OTT	Over-The-Top (unmanaged services)
PDF	Probability Density Function (also abused as Probability Mass Function)
QoE	Quality of Experience
σ	Standard Deviation
UHD	Ultra-High Definition
VoD	Video on Demand
VR	Virtual Reality

Bibliography & References

[1] E. Harstead and R. Sharpe: Forecasting of Access Network Bandwidth Demands for Aggregated Subscribers Using Monte Carlo Methods, IEEE Communications Magazine • March 2015, http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7060505&tp=?ALU=LU1005515

[2] BLC Mobility Report, https://pages.nokia.com/1503.bell-labs-mobility-report.html

[3] U.S. Census Bureau, https://www.census.gov/data/tables/time-series/demo/families/households.html

[4] U.S. Census 2015-2060 National Population Projections, https://www.census.gov/data/tables/2014/demo/popproj/2014-summary-tables.html

[5] Nielsen Q4 2016 Comparable Metrics report, http://www.nielsen.com/us/en/insights/reports/2017/the-comparable-metrics-report-q4-2016.html

[6] Nielsen 2015 Audience report, http://www.nielsen.com/us/en/insights/reports/2015/the-total-audience-report-q2-2015.html

[7] http://www.nielsen.com/us/en/insights/news/2010/u-s-homes-add-even-more-tv-sets-in-2010.html

[8] Massive VR or AR adoption will likely not occur until power consumption and battery storage allows for affordable, comfortable goggle form factors.
[9] Consumer Technology Association (CTA), 19th Annual Consumer Technology Ownership and Market Potential Study, via https://www.cta.tech/News/Press-Releases/2017/May/A-Smartphone-Surprise-U-S-Ownership-Hits-Record.aspx
[10] Tan et al.: Video Quality Evaluation Methodology and Verification Testing of HEVC Compression Performance, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 26, NO. 1, JANUARY 2016, pp. 76-90, http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7254155
[11] Purves D, Augustine GJ, Fitzpatrick D, et al., editors., Anatomical Distribution of Rods and Cones, Neuroscience. 2nd edition., National Center for Biotechnology Information, U.S. National Library of Medicine, https://www.ncbi.nlm.nih.gov/books/NBK10848/
[12] How Virtual Reality Hype Exceeds Demand—For Now, Fortune, http://fortune.com/2016/10/13/virtual-reality-headsets/
[13] The web is doom, https://mobiforge.com/research-analysis/the-web-is-doom
[14] How Fast Should A Website Load in 2017?, https://www.hobo-web.co.uk/your-website-design-should-load-in-4-seconds