Telzed

Calculating customer numbers and system capacity in broadband networks

A guide to basic network modelling

February 2022

File reference: Broadband capacity needed 24022022.docx

Details: A discusson paper to assist with the understanding of broadband networks

Author: R Steele

No liability is accepted for the accuracy of information contained in this report or for any decisions made or actions taken by any party, as a result of this report.

This report need not reflect the views of any client of Telzed.

This paper may include some comments and opinions that are designed to provoke additional thought and discussion.

Telzed Limited UK. Consulting, Advisory and Interim Management Services

All correspondence: Lilac Cottage, Westerfield Road, Westerfield, Ipswich, IP6 9AJ, UK. Tel: +44 (0) 777 178 7607

Registered Office: Orchard House Park Lane, Reigate, RH3 8JX, UK

Table of contents

Document history Ver 24022022 Initial version

1 Some network modelling methods are better than others

1.1 Background

Many strategists, consultants, modellers, business managers and engineers (and others) need to understand both how big (what capacity) and how many systems are needed. This defines the costs.

The system capacity of network systems like routers or mobile masts is finite. So a capacity limit defines the maximum number of customers and the maximum traffic that can be carried.

This is a well-known problem and solutions date back to the early days of telecoms with manual switchboards and early voice switches.

Most traffic is now packet switched. A customer gets a number of packets of data in a short stream, then they get nothing, then more packets. Each packet can be different sizes. Many packets, possibly almost continuously, are needed for a streaming service. This traffic usage needs to be understood from the basic fact that a user consumes a number of Gbyte in a month. This in turn defines an average Mbit/s ([Gbytes x 8]/([time x1000]). This Mbit/s usage must fit into a mobile mast (cell site) or other *point of concentration*. Traffic combines into a finite capacity-limited system – the network point of concentration.

The standard solutions and understanding are well known. But there are alternative approaches to the calculations and some of these do not always agree with standard network modelling, yet these alternatives persistently crop up. They *can* provide sensible results. But the alternative approaches could also cause *major errors* and so may over or under-predict costs significantly.

1.2 Different network modelling methods are possible

This paper provides a discussion of the methods used and how they can be applied. This is not intended to be complete, and network design engineers will use more exacting methods to build the broadband fixed and mobile networks that we all enjoy. These methods still relate to the simpler approaches described here. This paper is for the general telecom specialists, not for expert telecom engineers and network designers. Managers *et al* need to understand how customer demands and traffic impact network design as this defines investments, and so the revenues and margins. All decision makers therefore should appreciate the modelling methods used to provide the data on which decisions are made. Without this there are real possibilities that an incorrect modelling tool is used, with the inevitable bad consequences. Else the solidity of numbers is not properly understood, also leading to possibly incorrect decisions.

The methods in this document are sufficient for many modellers and they provide an essential platform for many economic or strategic insights. The methods have been adapted into real regulatory cost models that are used for decisions such as price controls. These models have been developed by this paper's author, and many other consulting firms, over c20 years. The methods are well proven. They *are* able to accurately model reality. The modelling defines the actual numbers of systems such as how many mobile masts are needed in a country or by a single Mobile Network Operator.

Three modelling methods are explained and demonstrated. This provides a foundation of understanding that should benefit many readers, especially “beginners” in telco modelling and cost/volume/revenue analysis.

2 Modelling networks - define the numbers of customers that can be served

A basic problem has existed for over 100 years. How many systems are needed and how big must they be? This defines the network costs.

If the system numbers *are* known (usually they exist or are defined in a business plan) then how many customers can then be served? How does this vary with the customers' usage of the service? The customer numbers and usage define the revenues, given a price per customer and price for traffic.

If revenue > costs then this is normally a good outcome! So, the calculations are at the basis of most business models.

We discuss three methods:

- **Traffic based calculations**. Customers consume traffic, measured in Gbyte. This defines a usage in Mbit/s. This usage depends on the time of day and variations by days of the week. The peak usage is in the busy hour. A network must be able to carry the peak (busy hour) demand, so engineers/managers make it big enough to carry this traffic. A complexity is that traffic is statistical – sometimes users download a lot or just a small email. If everyone needs a lot of traffic at the same time the physical capacity may be exceeded. This probability is always finite[1], but can be set at a tolerably low value. This traffic-based method can be termed a Gaussian-statistics based method, even if practical calculations do not need such mathematics.
- **Circuit or speed-based calculations**. In this model a customer is assumed to get a service at a defined speed (or above a minimum speed). This is similar to the "old world" of voice calls being switched circuits. These circuits from many customers making a call have to fit into the capacity of a voice switch or core transmission link. Both of which have finite capacity. This may be termed an "Erlang based" or circuit based method. It is adapted here for broadband.
- **Oversell factors or contention ratios.** These are easy to understand and are sometimes used for marketing or sales purposes. In some cases they can give similar answers to the above two engineering methods. Unless carefully defined and carefully used, they can give major errors, so are not normally used in network design (at least within most teams Telzed has dealt with). Overselling is also termed an over-booking factor (OBF). The contention term means that N services have traffic contending for the finite capacity. So 20 services can contend for every one service's worth of capacity - an overbooking of x20. It can also be termed as a "simplistic rule"

[1] Non-blocking systems can be used, but this is a specialised area and would worse case a network design – every customer using the maximum possible traffic simultaneously can all still get a clear service. This is normally too expensive as this situation is so rare, that it can be ignored. Why build a network so big that most of it is almost never all used? Some readers might look at dealer board systems from the days of switched voice calls. Non-blocking networks do exist, but are not covered here

or even a “make-up method” as it is easy to make up a factor that has little relevance to reality.

The problems to be addressed by any method are:

- Define the dimensions needed of a system (mobile mast or a fixed broadband router[2] for example), given the known usage by customers
- Define how many customers can be served, given the known limits of the mast or router
- What probability is there that the service cannot be provided due to the mast or router being over-loaded?

Once known, it is simple to define the costs, customer numbers and traffic so revenues and margins may be determined.

We consider each method in the following section.

[2] Satellites and many other systems are also governed by similar mathematical and engineering factors. See also the Satellite paper on the Telzed site. NB there are many relevant papers on this site that use the traffic-method defined here. The web site also advertises a universal mobile model – that uses the traffic methods discussed here

3 Examples of the three methods

In this section we examine the three methods. A simple-calculation approach is taken so that any reader is able to do the same analysis with only a pocket calculator or simple analytical tools (few lines of Excel or erlang tables). This simple approach is adequate to appreciate the methods and to understand where each may be appropriate or not.

3.1 Traffic based calculation

This method is the simplest and easiest to understand. It is also essentially what most network models actually use (with some variations). It reflects both real engineering practice and network reality. It is based around gaussian statistics, though there is little need to actually use gaussian formulae.

The basic inputs are:

- Number of customers N using the one mast or core broadband router
- The traffic per customer. Best understood as #Gbyte per month (or per year etc). Mobile usage is c10Gbyte per month. Fixed line broadband usage is c400Gbyte per month per customer. This varies internationally. These are used as examples. The same maths applies to (say) Fixed Wireless Access masts using perhaps 100Gbyte per month
- The system capacity. The Mbit/s of the router or mast. These have a finite limit. More traffic than this has to deleted or stored and must be sent again or later. We use examples of 200Mbit/s for a mobile mast and 10,000Mbit/s for a core fixed broadband router
- Time of day and day of week variations. A monthly average usage can be measured, estimated or predicted. The system has to cope with the peak demand in any one day (the most used day of the month). This varies by country. With 30 days and similar profiles each day and c13.5% of a day's worth of traffic in the busy hour[3], then the traffic in Mbit/s = 0.01 x #GByte per month. This comes from: [1exp9bits x 8 bits in a byte x 13.5%] /[30 days x 3600 seconds in the hour). The 13.5% traffic in the busy hour is country/network and culture specific[4]. The value used here is reasonable, and in line with some countries. It will vary over time as customer behaviour changes. Modellers should adjust this calculation to the local situation.

This means that 10Gbyte per month uses *on average* 0.1Mbit/s in the busy hour. 400Gbytes uses 4Mbit/s on average. This seems very low as a customer sees a service of say 100Mbit/s from the fixed line or over 20Mbit/s on a mobile device. The reason for this is that

[3] Most networks have a peak traffic period and are designed to cope with this peak. An hour is commonly used. Other periods may be used with minor adjustment to the calculations in this paper, but the results will be similar

[4] Time of day traffic profiles are easy to measure by network managers and have been published by some operators. Network managers monitor this data. This enables the percentage of a day's traffic in the busy hour to be calculated. Likely to be c9-15%

the service streams to the customer in bursts of 20Mbit/s or 100Mbit/s, then nothing is downloaded briefly then more bits are downloaded. Other users get to use the same capacity when the user is not downloading. A customer never uses 20Mbit/s continuously, it is used only briefly, then again later. The service only slows down if a lot of customers get their burst of 20Mbit/s or 100Mbits and are trying to do it at the same time. This happens, but unless the system is close to maximum usage, then the resulting slow down in service is small. It is rare for short bursts of traffic to coincide with many other customers. Slower downloads *are seen* sometimes in the busy hour, especially in mobile networks. This happens when many customers download at the same time.

If a normal mobile speed is 20Mbit/s, then it would need more than 10 customers streaming a large download at the same time for a slow down to be seen. The demand then exceeds the 200Mbit/s limit of the mast. Since each customer uses *on average* only 0.1Mbit/s this is a relatively rare situation, unless there are a lot of customers on the one mast. We can now calculate the value of "a lot."

Using the above Telzed rule of thumb, where traffic per customer = 0.01 x #Gbyte per month, the total busy hour traffic in Mbit/s from N customers is (N x 0.01x10) Mbit/s in mobile (=Nx0.1 Mbit/s) on mobile. The traffic is Nx4Mbit/s in the fixed line broadband system.

This means that 2000 mobile customers are possible on each mast (200Mbit/s / 0.1 Mbit/s per customer) or 4500 on each router (10,000/4). *These are ideal maximum values.* This approach has been used in Telzed papers to keep the calculation simple and to provide an optimistic capacity for networks.

Reality must account for:

- Customers vary in their use. Some use more, some less. There are statistical variances. This can be modelled as gaussian variances.
- Customers are not totally statistically independent. So a news update or new game released, will cause collective actions, so many customers increase their usage on "freak days".
- Engineers know that if the 100% limit is exceeded, customers complain and some systems need managed to avoid a packet overload "heart attack." Packets must be dropped or delayed. Customer systems then re-send the same lost packets (causing even more traffic) and so overloads can rapidly reduce service quality.
- Traffic grows at c30% per year [country dependent and differs fixed from mobile], so there is no point having all systems close to the limit as more capacity has to be added in a few months. Systems are best built for 6 or 12+ months of growth without technicians doing upgrades. Why send teams to a mast every month?

These traffic variances are not all random gaussian type. The net result is that most systems are run at c65%-80% of the absolute limit depending on the business and system in question. This is the *practical average maximum realistic usage rate*. Of course actual usage will vary by individual system/mast. Beyond this limit, more traffic will need investment in a new router, new mast (or new cell on same mast) or more capacity in the existing mast/system. Some systems when newly installed will be only at c30% of the capacity limit. An 80% figure risks more system overloads on the freak demand days in the busy hour, than a system with traffic averaging 70% of the physical limit.

Note that more traffic usage increases revenues per system. Business and engineering managers set the figures trading off revenue $s for service quality and the required investment.

Using a 70% figure, the 200Mbit/s mast can therefore "only" have 1400 customers. NB this customer-number per mast is similar to that seen in some developed mobile economies, where the average mast capacity is around this value.

Aside: few masts have the fabled 1Gbit/s capacity of 5G. Some old masts have only 100Mbit/s or less capacity if 2G or 3G based, or they cover low customer density areas.

Gaussian statistics governs the essential variance in traffic per customer in the busy hour or the number of customers downloading. The variance is roughly SQRT(N). So if a network designer or modeller wants the probability of overload to be less than X%, the standard gaussian formulae could be used. 3 variances (3 sigma) would mean 0.13% chance of overload. Then with 3 variances the maximum average traffic would be 187Mbit/s (93% fill factor or 1870 customers[5]). This is more than the usage limits shown above. This is because real networks (say with 70% design limit) are designed to allow for greater variations as the traffic variances are not truly random gaussian – as discussed above.

3.2 Circuit-based or Erlang-based modelling

The erlang or circuit-based approach considers that a customer gets a service at a defined speed. In effect the mast or router provides a "circuit" to the customer at say 10Mbit/s or 20Mbit/s (mobile) or at 100Mbit/s for a fixed broadband. Examples here are for illustration.

A 200Mbit/s mast can provide 20 off 10Mbit/s circuits. This does not mean only 20 customers. They would all be guaranteed a 10Mbit/s service. A perfect, never-blocked service. But they will almost never all use the service at the same minute or second. So more customers can be allowed. How many? This problem was solved by Mr Erlang.

The quality of service is defined by the tolerable chance of more than 20 customers making a download. One customer is then blocked or bumped off, or else all customers get less than 10Mbit/s. A reasonable value for illustration is 10%. A 10% chance or less than 10Mbit/s is the quality of service or blocking factor. 1% is a low value for broadband modelling purposes.

---

[5] This is not worth making much effort to reproduce, but if x is the average customer numbers, then 2000 - 3xSQRT(x) -x = 0. The 2000 value defines the customer numbers (and hence the traffic) with zero variance. This assumes the variable is the number of effective users. This is simple to solve. It is not worth effort as network managers know such factors and the real statistics and so system limits are based on other design rules (see above) and these lead to between 65% and 85%, as the allowable average usage of a system. A number of messages result: larger systems (more users) allow higher fill factors, as the variance is smaller (varies as SQRT). This also means slicing up a finite spectrum/mast or router capacity into say 5 services each with finite capacity allows LESS traffic in total than if they all share the same capacity. There are some reasons to do this, but it is less efficient from a network capacity view. The same is seen in erlangs – a system with 10 circuits allows more than 5x the traffic of a 2 circuit system. But once circuit numbers are very large, then the relationships become more linear. **Network slicing allows less traffic in total.** This can be overcome by using priorities and network management adjustments so that some services can use other services' capacity if their slice is overloaded. But this undermines the principle of dedicating fixed capacity slices to different services – it is not pure slicing. Why bother in the first place?

.

Next a user must refer to erlang tables[6]:

- There are 20 circuits available
- The bocking probability is 10%
- The resulting traffic is then 12 erlangs.

Two factors must always be known to calculate the third.

The method now needs to convert the traffic limit (12 erlangs) to Mbit/s. A single user consumes 0.1 Mbit/s on average. This is 0.1/10Mbit/s of a circuit, so the customers each use on average 0.01 erlangs. With a total capacity of 12 erlangs then 1200 customers are possible. This is not much less that the “correct” 1400 customer value defined earlier using traffic/gaussian methods.

The approach used here uses slightly different thinking to voice circuits and erlangs. The erlang value defines the occupancy of a circuit. It has no dimensions. One customer occupies a circuit for a fraction of the busy hour - the erlangs per customer. So usage of 0.1Mbit/s on average over the hour means an erlang occupancy of the 10Mbit/s circuit of 0.01 erlangs.

NB, a user has to adjust the calculation as the same traffic (0.1Mbit/s) is creating only 0.005 erlangs occupancy of a 20Mbit/s circuit. This adjustment is easy to overlook (not used in voice switch calculations).

If customers are to get 20Mbit/s as a service speed. This means 10 circuits or 10 simultaneous services. This allows only 4.45 erlangs in total (from tables). The maximum customer numbers is therefore 890 (4.25 erlangs/0.005 erlangs per customer).

Note that mobile masts do not normally work with specified-circuit speeds. Customers get what ever is available. Maybe up to 200Mbit/s if no other customers use the mast. At the very busiest period they may briefly get just c0.1Mbit/s but most download bursts will be far faster. One data-burst’s speed will vary from another – they are not set at 10Mbit/s. **However the erlang method is sometimes reasonable and it gives an approximation of the customer numbers when a modeller wants to think about service speed.** Care is needed to define the erlang value per customer, as that changes with the notional circuit (service) speed.

The fixed line 10,000Mbit/s router has 100 off 100Mbit/s circuits. Or capacity for 84.05 erlangs using 10% blocking factor. The erlangs per customer are 4Mbit/s/100Mbit/s =0.04. The router can therefore have 2,101 customers. The traffic method with 70% fill factor of the router allows (10,000 x 70%)/4 = 1750 customers. So the values (2101 and 1750 are similar). This is expected as erlang and gaussian methods converge for large traffic values.

The erlang result of 1200 mobile customers is not far from the earlier engineering-adjusted gaussian traffic value of 1400 customers. This means that erlang methods can give sensible results, *if* a user considers the relevant factors properly. However masts and routers do not normally work by providing fixed speed circuits.

[6] E,g. https://www.erlang.com/calculator/erlb/

Although thinking about a customer's service speed may seem easier to understand, it is recommended to use the simple traffic method. No erlang tables are needed and it is easy to adjust to different network management views on how full systems can be allowed. An aggressive CFO may demand usage limits over 80% before more expansion funds are released. Lower fill % values increase quality and gives longer times before additional expense is needed to enlarge systems. But high fill factors increase traffic and $ returns per system. Network managers and designers already know the system fill factors, traffic profiles, trends and system limits[7].

3.3 Oversell factors

It is possible to decide that every customer's service is oversold (or the network is over-booked). So if customers are to get 10Mbit/s (or 20Mbit/s), and the mast has 200Mbit/s capacity then the customer numbers are 20 (or 10), with no oversell. With an oversell factor of 20 then you have 400 (200) customers on the same mast. Or 1000 (500) with an oversell of 50.

In the fixed line case there are 100 off 100Mbit/s services possible with no oversell. So an oversell of 50 enables 5000 customers.

So, you can have any number of customers depending on the chosen oversell factor. The factor gives no indication of the service quality – the network could be very under used (high quality) or overloaded. Using the values above, most results have nothing in common with the customers calculated using the traffic or erlang/circuit methods.

The deep question is how to define the oversell factor. Just choosing a value of (say) 20 is best referred to as *making it up*. It sounds good to customers and makes a network simple to design. It is mostly useless.

If we use the realistic network values of 1750 fixed line customers, then the real oversell factor is 17.5 in fixed. Using 1400 mobile customers, the real mobile oversell factors are 140 if 20Mbit/s is the notional service speed or 70 if the service is assumed to have 10Mbit/s services. This is the actual service numbers divided by #servcies with no oversell.

It should be clear that almost any customer number or oversell factor can be defined. One can be traded for another. But there is no link to the actual traffic.

A 50 oversell factor is hugely different from the actual 17.5 in the fixed line case.

10, 20 or 50 oversell factors for a mobile model is totally unrelated to the "real" values. Even the oversell values of 70 or 140 are not realistic anyway as *mobile masts are not used as fixed speed circuits*. The values have no similarity to the 10, 20 or 50 oversell factors used in some (incorrect!) modelling, as design rules.

Using notional oversell values, the chances of correct modelling results aligning with calculated values based on traffic theory or on erlang theory, is slim.

---

[7] Non-technical managers may discuss their modelling with them, if needed to supplement this paper. Some modellers/managers may think they know better. Else: confirm this Telzed paper with other consultants who have the right background. Or think and use the example formulae provided here

3.4 Summary of the alternative methods

The following messages may be taken:

- Traffic analysis using conversions to busy hour, does model real networks, if realistic fill factors for systems are defined (likely 65-80%). This is a variation on gaussian statistics, but adapted to reflect real network traffic variances, that must allow for freak days and traffic growth. Traffic does not strictly follow gaussian statistics. The calculations are trivial pocket calculator type, but a user still needs an understanding of engineering limits, and customer traffic demands

- Erlang methods can give reasonable values. They can be close to the traffic analysis (real world) calculations. Most systems do not provide switched circuits with specified speed to customers, so the method does not exactly reflect most packet-based networks. It is ideal for switched voice circuit networks, or similar, but *it is not ideal for modern broadband*

- Oversell factors are a poor method to model a network. The rule is simple to apply but can give hugely wrong results as traffic usage is not considered. *Traffic* drives network costs. This traffic varies over time and by network. Yet, if using oversell, the network dimensions do not change, as that is set by: oversell factor x customers' notional speed.

4 Conclusions

This paper shows the need for all telecom experts to understand some traffic theory and elementary modelling. This enables business plans, network models, cost models and demand models to be understood and combined.

The best method uses traffic modelling (aka modified gaussian thinking). It is both simple and accurate if real engineering limits of systems' capacities are used. The practical limit is less than 100% of the physical limit. Network or business modellers and managers must understand how *traffic* (#Gbyte) relates to busy hour *usage* (in Mbit/s) based on time of day demands and how traffic varies over time. The method has been verified in models that calibrate to actual network numbers (such as numbers of required masts).

Readers who need to learn more can engage consultants or discuss the problems with technical departments in a telco. They should all use similar thinking. Or contact Telzed.

The issue is profound and proper understanding of the demand and how this varies over time and by time of day, explains why most telecom networks do not overload regularly in peak demands. It also shows why home working with Covid did not cause network overload. That was predicted by many who did not understand that the busy hour is in the early evening (most countries) and home working increases *day time* traffic. Telco engineering managers did publish time of day data in 2021 and stated they had designed-in headroom for growth and peak (freak) variations. They also understood the low impact of day-time traffic. Using this paper, readers will understand the essential principles that these managers use and readers may wonder why some dismissed their claims, predicting network meltdown.

The paper shows that any model basing the design on oversell/overbooking/contention factors must be examined very carefully. It might be right, but it is possibly almost useless.

Please contact Telzed for further advice and help

See Telzed web site for additional papers


Telzed