Connecting the Last Billion is a Solved Problem

Micah Beck
Min Kao Dept of Electrical Eng. and Computer Sci
University of Tennessee
Knoxville, Tennessee, USA
mbeck@utk.edu

Abstract— The problem of serving the critical digital connectivity needs of every person in the world is a solved problem. The key challenges of achieving critical mass and bringing down the costs to a level affordable even in very low income countries and communities can be met with current technology. Unfortunately, the solution is not acceptable to the telecommunications industry or its surrogates in government and community organizations. Thus we continue to pursue expensive strategies that are difficult to deploy globally but which conform to the assumptions and business models of the telecommunications industry. Those efforts may benefit the next billion unconnected individuals, but it will likely never reach the last billion --- those in the most remote, disrupted and economically disadvantaged communities, whose connectivity is delayed ad infinitum.

Keywords—universal broadband, caching, asynchronous communication, content delivery, latency hiding, edge processing

I. Introduction

In this paper we address these specific issues that are raised in the symposium Call for Papers:

> “There are numerous projects and initiatives ongoing around the world, however, these are fragmented and lack the critical mass and coordination to be able to impact the future standards, product development, and cost of deployment otherwise achievable by volume.
>
> While technologies are available today, they need to be customized and optimized at the systems level to bring down the cost of the network in order to be affordable.”

The problem of serving the critical digital connectivity needs of every person in the world is a solved problem. The key challenges of achieving critical mass and bringing down the costs to a level affordable even in very low income countries and communities can be met with current technology. Unfortunately, the solution is not acceptable to the telecommunications industry or its surrogates in government and community organizations. Thus we continue to pursue expensive strategies that are difficult to deploy globally but which conform to the assumptions and business models of the telecommunications industry. Those efforts may benefit the next billion unconnected individuals, but it will likely never reach the last billion --- those in the most remote, disrupted and economically disadvantaged communities, whose connectivity is delayed ad infinitum.

In this paper we focus on the critical applications which are the means by which connectivity delivers value to end users, rather than on the service of connectivity itself. We do not begin by assuming some particular form of connectivity service as the goal, as is typical among proponents of “high quality” universal broadband service[1]. If we made such an initial assumption, then we would not be able to evaluate the appropriateness of that service as a tool for benefiting end users.

In order to explain the claim that the problem of critical connectivity has been solved, we must consider what critical applications are enabled by digital connectivity. Typical critical applications that are enabled by digital connectivity include:

- Remote education and healthcare
- Applying for government aid, seeking permits and obtaining public information
- Discovering the availability and pricing of goods and placing orders

[1] “Today, the FCC recognizes high-speed Internet as the 21st Century’s essential communications technology, and is working to make broadband as ubiquitous as voice, while continuing to support voice service” [1] (https://www.fcc.gov/general/universal-service).

It is worth noting that none of these crucial applications need to be highly interactive, and so do not require low latency or constant availability in network connectivity. They can all be served quite adequately using forms of connectivity that have very weak bounds on latency and on the duration and frequency of service interruptions, if average bandwidth is sufficient.

It is generally understood that strengthening the network performance requirements for an application, for instance minimumizing bandwidth or maximumizing jitter, can reduce the potential population that can be served. The converse is also true - if network performance requirements are weakened, then the potential population that can be served is increased. This is consistent with a general principle of layered systems, the Hourglass Theorem [Beck99].

- Evolving an application to take advantage of stronger assumptions than its original design is generally straightforward. This is because no assumptions made by the current implementation are violated. Thus an application can be used “out of the box” under stronger network assumptions, and new versions can then be improved with new features which take advantage of them at the convenience of the developer.
- In contrast, when network performance assumptions are weakened, the current implementation of an application may degrade in its value to the end user. In some cases, the application may cease to function altogether if key implementation assumptions are not met. The explanation among developers and user alike is typically that a network with weaker assumptions is inadequate and“cannot support the application”.

The fallacy in this latter point of view is that the failure of an implementation developed under strong network performance assumptions to operate correctly under weak assumption does not mean that the application *cannot* be implemented. It may mean that the implementation may have to be modified or, in some cases, redesigned. The application interface may change to adapt to weaker network performance assumptions. Only in extreme cases can the end user’s goals in using the application *impossible* to achieve under weakened assumptions.

Some specific classes of applications, such as telepresence and real time control, do have requirements that actually cannot be implemented under weak assumptions. But for the others, some additional investment may have to be made in programs and perhaps in edge resources such as data caches to meet user requirements under weaker assumptions. This analysis explains why the application development community and the end user community tend to resist the notion that larger populations can be served by weakening the minimum network performance assumptions under which critical applications must operate.

Most of the most widely used and most important classes of application can be implemented under assumptions that allow for high network latency and intermittent connectivity. This is especially true when high average bandwidth and the availability of edge storage resources can be used to hide or ameliorate the impact of these network characteristics from end users [Beck and Moore, 2022]:

- Any service that can be implemented through the asynchronous exchange of messages, forms, media clips and other digital objects can be deployed under very weak network performance assumptions.
- Browsing through directories of possible content and services can be supported through edge caching and limited disconnected interactivity.
- Streaming media can be delivered in large chunks and buffered in edge storage, either as on-demand streams with initial delay, as subscriptions declared by the end user, through learning algorithms or specified by policy, or through less synchronous request/response.
- While remote in-person tutoring and medical office visits have strong network performance requirements, many other important types of remote education, healthcare, e-government and e-commerce do not.

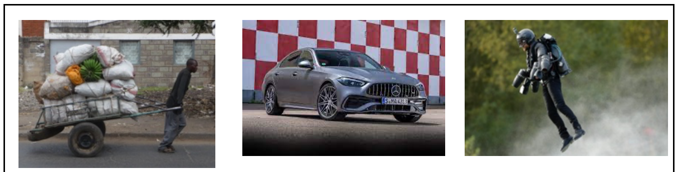
Let us undertake a thought experiment, and ask what the implications might be if all possible critical applications were implemented under very weak assumptions of network latency and connectivity.

- Everyone on the planet who could be equipped with a client no more powerful than a low end smartphone would be able to use critical network applications at little or no recurring fee for connectivity.
- Forms of connectivity with high latency and intermittent service (e.g. satellite, aerial transponders, storage carried on vehicles) would be capable of delivering critical applications to otherwise disconnected communities.

- Very limited edge resources such as data caches and in-browser programmability could be used to provide constantly-available services to end users, hiding the delays imposed by network connectivity constraints.

This thought experiment is a nightmare for the telecommunications industry, which is fully invested in the idea that adequate connectivity can only be provided by expensive "high quality broadband" service. Advocates for the "*next billion*" individuals to be connected also recoil at the idea. They assert that it is absolutely necessary to subsidize free voice, teleconferencing and remote gaming services in order to serve those who can afford connectivity the least. The *next billion* represents an emerging market for profitable services delivered over networks with high performance requirements. So providing them with high quality broadband is a good short-term investment which results in no disruption of the application or edge resources markets. The *next billion* speak with a voice that is not as loud as the first or second billion that were connected, and their pockets are not as full. But their voice is much louder and their pockets are much more full than the "*last billion*" who must wait to be connected at some point in the indefinite future.

An analogy might help to illustrate the injustice of the situation. Consider one of the many workers who delivers goods in rural Kenya by pushing a mkokoteni, or human-powered cart. He has saved up to rent a Tuk-tuk, or three-wheeled motorcycle that has a cargo area, but there is no petrol station nearby. The government would pay to create a station and provide standard grade unleaded petrol. But his neighbor who wants to buy a Mercedes insists that the station should only be built if it will serve high octane fuel. He claims that it would be insulting to limit the community to standard grade petrol. Someone else heard that Elon Musk predicts that everyone will soon be commuting by personal jetpack, and so they think that the station absolutely must sell jet fuel. In the meantime the man continues pushing his mkokoteni every day.



The last billion | The next billion | Tomorrowland

Figure 1: Who should the petrol station serve?

Given the forces standing in the way of the adoption of weak latency and intermittency standards, how would we be able to establish them? Here is a plan which proceeds step-by-step:

- Propose a standard set of minimal, weak assumptions under which network applications must be able to adequately serve end users.
- Develop a set of application requirements that meet end user needs while being implementable within those standard minimum assumptions (SMAs).
- Develop a set of prototype applications demonstrating the ability to meet those application requirements under the SMAs, revising those assumptions if necessary.
- Advocate in standards bodies and public service providers for the adoption of the SMAs as a standard for network accessibility of all critical services.
- Advocate with infrastructure policy makers and operators for the value of deploying networks which, at least initially, may minimally satisfy the SMAs. These should ideally be usable at little or no recurring connectivity cost to the end user.
- Work with the development community to enhance the quality and variety of applications available under the SMAs.
- Work with edge infrastructure providers to provide affordable edge resources such as generic data caches which can enhance the quality of end user experience in applications deployed under the SMAs.

Note that this strategy in no way precludes the construction of infrastructure that provides service stronger than the SMAs require. Nor does it prevent application developers from building applications that work better when stronger infrastructure is available. It simply allows the community to benefit from the early deployment of critical applications which current cannot be delivered over universally deployable infrastructure. This may require some additional effort in the evolution of high end applications if they want to also qualify to fill roles that are critical to the public.

Implementing critical applications to run on infrastructures that satisfy such an SMA would immediately benefit those who would otherwise be relegated to the "*last billion*" to be connected sometime in the indefinite future. But it would also benefit every end user who even occasionally has to function in an environment in which strong connectivity assumptions do not apply. Which is everyone. [4]

The telecommunications industry has developed a strategy of marketing connectivity services as if the strongest claims they can legally make always apply. A well-understood example of this is the marketing of last-mile connectivity according to the highest aggregate bandwidth that it can possibly deliver. The actual throughput experienced by any particular user is dependent on a variety of factors, including congestion in the core network, the capacity and distance of the endpoint being reached, the error rate, transport and user level protocols employed at endpoints, the memory and processor loads on the endpoints and a variety of application-specific considerations. The total aggregate bandwidth actually utilized by a business or residence and the impact of the bandwidth capacity of the last-mile connection is unknown to the end user or the policy maker.

Less well understood are the variability of network error, high latency and frequency of disconnection suffered by end users. These characteristics are not advertised to end users and no particular bounds are included in the definition of “high quality broadband service”. These quality of service characteristics are impacted by factors such as the topology of the edge network (often wireless) in relationship to the area it serves, the age and condition of edge routing equipment and terminal devices, wireless interference and wire cuts, natural disaster, investment levels and technical competence of the ISP and end user mobility. Marketing and lobbying by the telecommunications industry promises that each new generation of transmission technology and end user equipment will eliminate all such factors, delivering a world in which connectivity can be relied on for the most critical of real-time activities. The truth is that today even a connection to a Cloud server delivering routine office-automation services such as word processing and data visualization is liable to be degraded or interrupted at an inopportune moment.

Applications and edge infrastructure capable of connecting the last billion will also be resilient to the real-world challenges that plague network users even in communities that are already making huge investments in infrastructure. The techniques used to hide connectivity underinvestment or disruption will also hide the gap between marketing and reality. A user who finds themselves temporarily disconnected by a storm will be able to continue critical tasks. Wildfire fighters in remote areas and warfighters on distance battlefields will also have the critical services they need to perform their vital roles. In times of network congestion certain critical functions will remain usable. Vehicles traveling on rural highways need never experience service blackouts that make critical services completely unavailable. [5]

But what of the people who currently have to choose between connectivity to enable remote education for their children and food to feed those children? How will these people connect to the immersive metaverse or engage in high volume stock trading? Should they be provided, at little or no cost, an edge network that does not support Cloud connections with 10ms latency? Is it better that they have no connectivity until their government sees fit to subsidize high quality broadband networks for them? It is true that if people can use a cheap network for the critical needs that enable their survival and engagement as digital citizens, their activities may be less easily monetized. But is this a bad thing? Exploiting end users as a source of wealth is seen as necessary and desirable by many who advocate for universal broadband service.

Maintaining strong minimum assumptions about the network may help ensure that the next billion to be connected have the best service that they (or their government) can afford. But it has the effect of delaying the connectivity of the last billion, perhaps forever. A policy that required every end user to be provided with free or affordable SMA-compliant connectivity, while allowing higher levels of service to be provided at a reasonable price would have an unknown effect on the market for high-end network services. It is argued by broadband advocates that it would result in many more end users never having higher level service available to them. An alternative view is that it would expand the market for networking by enabling society to more completely rely on connectivity for critical applications without excluding those who cannot pay.

The SMA-based approach to universal connectivity is technically sound, affordable, low risk and could quickly alleviate the lack of connectivity to large numbers of those currently least able to pay for it. But it is widely viewed as unrealistic and perhaps even pernicious by the global connectivity industry and its policy advocates. Moving in this direction would require a significant commitment by the universal service community to put the critical needs of the *last billion* before the immediate commercial interests of telecommunication providers.

References

[1] FCC, “Universal Service,”, URL https://www.fcc.gov/general/universal-service, retrieved 8/21/2023

[2] M. Beck, Communications of the ACM, 62(7), 2019, pp48-57. ACM.

[3] M. Beck and T. Moore, , “Is Universal Broadband Service Imposible,” in IEEE 19th International Conference on Mobile Ad Hoc and Smart Systems (MASS 2022). Denver, CO. Oct 19-21, 2022

[4] highspeedinternet.com, “Universal Service,”, URL https://www.highspeedinternet.com/resources/why-does-my-internet-keep-disconnecting, retrieved 8/21/2023

[5] [5] Micah Beck, Nancy French, Ezra Kissel, Martin Swany, Terry Moore, “Data Logistics: Toolkit and Applications,” in GOODTECHS 2019 - 5th EAI International Conference on Smart Objects and Technologies for Social Good, Valencia, Spain, September