An end-user approach to an Internet Score

Christoph Paasch, Kristen McIntyre, Omer Shapira, Randall Meyer, Stuart Cheshire
Apple, Inc.

Introduction

Consider the Internet today. The end-users have become accustomed to measure their Internet in terms of throughput; the operators and the vendors have been marketing their products in terms of throughput; available network measurement tools are solely focusing on throughput. However, throughput is a very bad indicator for the quality of a network and the user-experience one can expect. There are countless other factors that drive the user-experience, like latency, responsiveness under working conditions, the ability for the network to allow modern protocols to be used, and so on…

Unfortunately, the Internet's ecosystem, constituted by the different vendors, the operators, the end-users and software, seems to have adopted throughput both as a primary quality indicator, and as a primary predictor of the expected user experience.

We believe that in order to encourage change, a new metric should be used as the quality indicator / experience predictor. Such metric should be based on the various fundamental properties of a network and endpoint stack that define the user experience. We further posit that such metric can be adopted to the different user types (from regular email/browsing to gamers to the professionals working from home) by weighting the above fundamental properties.

In this vision-paper we introduce such metric, which we call INTERNET SCORE and demonstrate how it can be defined to provide a fair and accurate evaluation of the network quality from the perspective of the end-users. We start by outlining some of the properties of a network that do influence the user-experience, and continue in the next section by identifying the characteristics that the INTERNET SCORE should respect. We formalize the definition of the INTERNET SCORE as a function that uses the network properties as input and provides a single scalar output that respects these characteristics. We finish this paper with some suggestions on next steps.

As this is a vision paper, we do not yet go into the details of how such INTERNET SCORE will transform these multiple inputs into a single score, but rather present this as a potential framework from which future collaboration can originate from.

Network Properties

A large range of network properties influences the user-experience. We can classify these properties in 4 different categories:

- GOODPUT: This measures the amount of useful data the device can send/receive within a unit of time.
- IDLE LATENCY: A measure of the transaction latency when the network is not actively being used.

- WORKING LATENCY: A measure of transaction latency when the network is being used. Our data shows that during active use, the latency may change dramatically. It is important to measure this to evaluate the network's ability to do multi-tasking (aka., execute large data-transfers while being responsive to short transactions).
- PROTOCOL CONFORMANCE: Various standardization organizations are continuously innovating the core Internet protocols, to allow higher throughput, lower latency, better security and privacy to improve the overall user-experience. Sometimes, networks are interfering with the deployment of those solutions. The property of a good network would be that it gracefully supports those technologies.

These properties define the overall characteristics of a network. However, another dimension is important. Namely, the means of communication used on the Internet. For example, network's goodput may be measured with a tool like **`iperf`**, or by using an **`HTTP GET`** operation. In the latter, there is even the choice of running the **`HTTP GET`** with **`HTTP/2`** over **`TLS`**, or **`HTTP/3`** over **`QUIC`**. Each of the above methods will yield different, and potentially mutually incompatible goodput values.

Similarly, there are various (and potentially conflicting) ways to measure the idle latency. While a naive approach would be to use the **`ping`** utility to measure the **`ICMP`** latency, results of such measurement are unlikely to be relevant to the end-user experience. Different IDLE LATENCY metrics, such as the **`DNS-request`** latency, or **`TCP-handshake`** latency are more suitable for the purpose of establishing the INTERNET SCORE, as those are typical transactions in the real world.

Considering that we want the INTERNET SCORE metric to quantify the network quality from the perspective of an end user, it is apparent that the network properties should be measured in the context of real-world use cases.

Because of that we are constraining the measurements to the protocols such as **`HTTP/2`**, **`QUIC`**, **`DNS`**, **`TLS`** and the like, which are relevant to the experience of the end-users. Tools like **`iperf`** or utilities such as **`ping`** or **`traceroute`** are not representative of the end-user experience.

With this, the network's properties and their measurements will result in a large set of distinct scores, each quantifying a specific aspect of the network quality. An example is given in the following:

Goodput	**Idle Latency**
Upload HTTP/1: 40Mbps HTTP/2: 12Mbps HTTP/3: 15Mbps ***Download*** HTTP/1: 231Mbps HTTP/2: 112Mbps HTTP/3: 109Mbps	TCP Handshake: 25ms HTTP/1 Request/Response: 10ms HTTP/2 Request/Response: 30ms HTTP/3 Request/Response: 29ms DNS-request: 11ms

Working Latency	Protocol Conformance
Latency on load-bearing connections: 506ms Latency on connections different than the load-bearing connections: 47ms	ECN-support: No IPv6-support: Yes Wi-Fi security: Yes

Table 1: Example of user-relevant network properties and possible values and units.

It quickly becomes clear that the INTERNET SCORE will be an aggregation of a very large number of continuous and discrete values. Some values are "the higher the better", some values are binary, …

The following section will describe the desired characteristics of the INTERNET SCORE metric, and how the aggregation of the different network properties into the INTERNET SCORE would look like.

Desired Characteristics of the Internet Score

The goal of the INTERNET SCORE is to present a notion of Network Quality to the end-user. As countless network properties have a direct impact on the user-experience and thus the perceived network quality to the end-user, this score is an aggregate of all of these network properties and thus presents the quality in an abstract way. As such, certain characteristics of the score should be respected so that it becomes intuitive to the end-user.

Since the INTERNET SCORE is trying to capture the ineffable notions of 'quality' and 'utility' for any given Internet connection, we represent it with a single positive dimensionless number.

As a measure of utility, the INTERNET SCORE value should reflect poor network performance via low values. For example, an incident in a major important area that renders the network nearly useless, and causes the idle latency to exceed 4 seconds, should yield a low INTERNET SCORE value. Similarly, exceptional utility should result in a high INTERNET SCORE value.

To ensure forward compatibility with the future technologies, the INTERNET SCORE value would be a positive real number with no upper bound, with higher values representing better quality and utility. This way, as the technology advances, new relevant metrics can be combined into the INTERNET SCORE without the need to re-introduce the score to the users.

Perception considerations

Given that the INTERNET SCORE is dimensionless, there is freedom to set the typical values within a range that allows easy comparison and is understandable by users. The target is to keep the range within 4 integer digits, or 10^4. While it is likely that the INTERNET SCORE value will be computed as a floating point number, it is desirable to present it as an integer. This implies that values below 1 should not contain useful information for presentation.

It is important to note that while there are certain user perception biases being applied, this INTERNET SCORE is not fundamentally psychometric (i.e. parameters that are not directly user-observable are used to compute the INTERNET SCORE). It is based on characteristics of the network and underlying protocols that should correlate well with user perception of performance, however, there may be other factors that might or might not affect the user experience. Making the INTERNET SCORE correlate well with user experience is a nuanced topic, however, that still requires some understanding of the underlying psychology of user-perceived performance.

Timescale considerations

The INTERNET SCORE value is envisioned to be a snapshot in time of the conditions as they are measured. While not entirely instantaneous, the intention would be to evaluate conditions over a relatively short timescale so that, when required, post hoc smoothing and filtering can be applied to reduce variance and do trend analysis. This avoids erasing information too early in the process and precluding future changes to how INTERNET SCORE data might be aggregated.

Forward-compatibility considerations

Advances in the underlying technology must be accounted for to make the INTERNET SCORE useful into the future. New protocols and technologies come and go as technology and research dictate.

A guiding principle is that as technology evolves and connectivity improves, the INTERNET SCORE for an optimal connection should only increase monotonically. This is a difficult principle to guarantee, as noted below. To the extent that it is realizable, an INTERNET SCORE value that increases with improvements avoids having to re-baseline periodically as technology and infrastructure evolve.

Let's use the **`QUIC`** protocol as an example to clarify the above point. The performance of **`QUIC`** depends, among other factors, on the **`UDP`** performance. Since **`QUIC`** is very relevant at the time of writing, the relative weight of **`UDP`** increases as well. In the future it is possible that **`QUIC`** could be replaced by an even better performing and more capable protocol. In that case, the value of **`QUIC`** support would likely diminish while the newer protocol is in ascendence. Depending on the nature of the future protocol, the relative weight of **`UDP`** may decrease as well.

Balancing these waxing and waining technologies such that we always achieve a monotonically increasing score with improvements in the overall network is a challenging if not impossible prospect. Careful consideration must be given to how changes are valued so that the INTERNET SCORE can appropriately reflect the underlying improvements. A mechanism is needed to allow the INTERNET SCORE to adjust to these changes, similar to how stock indexes like the **`S&P 500`** weight the composition of its basket of stocks to adjust to changes in the underlying businesses over time.

Suggested scoring mechanism

The proposed mechanism for allowing future adjustment is a kind of weighting table. Typically a weighting table contains scaling factors to give a multiplicative 'weight' to transform input values to output values for a given factor. In this case, since weights can be quite complex, a table of weighting functions is proposed. This enables nonlinear transfer functions to be inserted that can react to any number of input parameters. This allows a single point of adjustment for the perceived value of protocols and technologies. It also allows having multiple weighting tables to measure different aspects of network performance. For example, there might be a weighting table for gaming and other real-time collaborative activities that emphasizes low latency, while another table represents a user who is rather a heavy user of streaming video. This complicates the notion of a single INTERNET SCORE to measure network performance, but the reality is that different users have different requirements. Inevitably, aggregating a large number of parameters into a single value erases information that may be relevant. Slicing input parameters in a small number of different ways can resurface some of that erased information in a way that users can easily understand. A weighting table makes this possible.

How it all fits together

From the perspective of having the INTERNET SCORE present useful information to a user not steeped in the nuances and details of network protocols and performance, measured parameters must be transformed in ways that attempt to make INTERNET SCORE values meaningfully predict user experience. When human perception is introduced, many of these transformation functions become non-linear, as human perception is often non-linear.

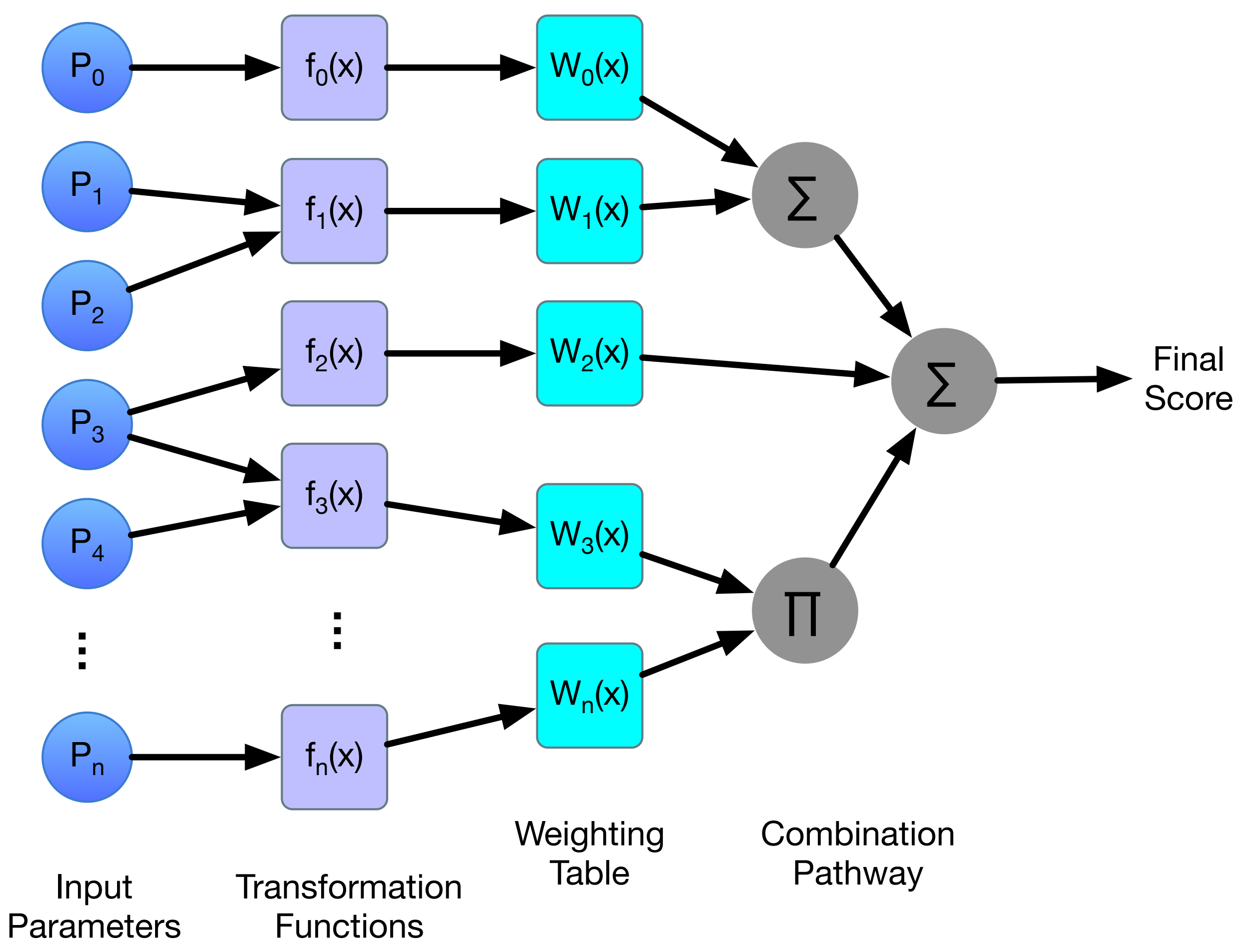

Figure 1: How the network properties from Table 1 can be used as Input Parameters to generate the final "Internet Score".

Examples of this abound (reference MOS scores, other ITU curves, A-weighting, μ-Law compression, reaction time studies, Fletcher-Munson curves, etc).

In principle, scoring systems of this nature can be represented by the block diagram presented in Figure 1. The various transformation functions that apply to one or more measured parameters are what define the ultimate characteristics of the INTERNET SCORE. Choosing appropriate functions for each of the input parameters is more of an art than any clearcut scientific process. Sometimes there are hints from our own life experience. Data from user studies can also be used to determine some curves to be fit and the coefficients that yield best correlation.

Some of the transformation functions are discontinuous *Heaviside step functions*. They typically represent the presence or absence of a capability in the network protocol stack. How the capability is weighted is a determination that is quite subjective, and is something that will be subject to the weighting table mentioned above.

Outputs from the transformation functions can be combined in various ways through the combination pathway. In some cases, where the output value is bounded, they can be combined multiplicatively, in a manner similar to how probabilities are multiplied. This is particularly relevant where the influence of those parameters is that they must all be present for the benefit to be seen - the benefit is AND-ed. Addition of outputs of the transformation blocks is appropriate where the benefits are relatively independent - the benefit is OR-ed. Outsized benefits can be accounted for in the transformation functions using appropriate mathematical techniques and re-scaling the values for later summation. In some cases, other non-linear combinations may be appropriate, like root-mean-squared, normalized vector distance, harmonic / arithmetic / geometric mean, etc. The appropriate use of these combination techniques on any given set of parameters has not yet been determined.

Next Steps

Producing an INTERNET SCORE that is both useful and an accurate reflection of the utility and performance of an Internet connection begins with building a model framework to explore how different transformation functions act on input parameters. The next steps toward a usable INTERNET SCORE will be to build this model framework, feed back its output to match the design parameters and goals specified above, and make both manual and automated adjustments. Synthesized parameters as well as real-world data will be used to cover as much of the parameter space as is practical and to press the boundaries to discovery any anomalous behavior.