Internet Society Pulse

Internet Resilience Index Methodology

2023

July 2023 v1.0

Table of Contents

Introduction

About the Index

The Internet plays a critical role in society today. The COVID-19 pandemic has further underlined the importance of reliable Internet connectivity for everyone. Unfortunately, not all countries are on a level playing field with regards to a resilient Internet infrastructure. Many low-income countries usually have under-provisioned networks and cable infrastructure, or they lack redundant interconnection systems. In these countries (or regions), the likelihood of Internet outages occurring is much higher than in other countries.

Measuring Internet resilience is not an easy task as there are several building blocks underpinning the Internet's complex infrastructure. Additionally, the Internet landscape varies considerably around the world and to be able to objectively compare countries between one another - on a common ground - there needs to be an objective set of metrics that tracks and records the different components that contribute to the resiliency of the Internet.

To achieve this task, the Internet Society is introducing the Pulse Internet Resilience Index (IRI). This document outlines the approach used to build the Index, the selection of indicators and the underlying data sources, the weighting scheme, and the aggregation method used.

The Four Pillars of a Resilient Internet Ecosystem

To grasp the multi-faceted nature of the Internet, the Index is built on four main pillars, which together contribute to the smooth operation of the Internet. The pillars are:

1. **Infrastructure**: The existence and availability of physical infrastructure that provides Internet connectivity.
2. **Performance**: The ability of the network to provide end-users with seamless and reliable access to Internet services.
3. **Security**: The ability of the network to resist intentional or unintentional disruptions through the adoption of security technologies and best practices.
4. **Market Readiness**: the ability of the market to self-regulate and provide affordable prices to end-users by maintaining a diverse and competitive market.

The Internet Society Pulse IRI is built using existing best practices according to the EU-JRC and the OECD Handbook on Constructing Composite Indicators[1] and uses the same methodology as currently existing indices such as the GSMA Mobile Connectivity Index[2], the Facebook/EIU Inclusive Internet Index[3] and the Web Foundation Web Index[4].

Data Sourcing

Selecting Indicators

Building a robust composite indicator requires careful selection of the underlying indicators. To date, there are no direct and readily available metrics that provide information about the resilience of a network or a country. In the Internet Society Pulse IRI framework, the indicators selected are reflective of a specific aspect of resilience that needs to be quantified. The OECD and EU-JRC handbook provides some guidance on the main characteristics to consider when selecting the indicators. In essence, they should be accurate, timely, and should cover as many countries as possible. Additionally, the Internet Society Pulse IRI relies exclusively on quantitative indicators as opposed to qualitative ones such as *perception of Service Quality*. This is to ensure that there is an objective set of metrics that can be used to make comparisons between countries.

Selection Criteria

The following criteria were used when selecting the datasets:

- Relevance: The indicator should work towards showing an increase or decline in the resilience of the Internet in a selected country.
- Accuracy: The indicator should correctly estimate or describe the quantities or characteristics they are designed to measure.
- Coverage: The data should cover as many countries as possible, as the Index is intended to be a global index. An indicator is not included if there is missing data on more than 25% of countries in the Index.
- Freshness: Any dataset should be at most two years old. Some datasets such as performance or network coverage should be recent. Some other datasets such as number of exits points do not change considerably over years, so it is acceptable to use a dataset which is a year or two old.

[1] https://knowledge4policy.ec.europa.eu/publication/handbook-constructing-composite-indicators-methodology-user-guide-0_en
[2] https://www.mobileconnectivityindex.com/
[3] https://theinclusiveinternet.eiu.com/
[4] https://thewebindex.org/

- **Continuity**: To objectively compare the index over the years, it is important to work with a stable list of indicators, which will provide data consistently over time.

Types of Indicators

There are three main types of indicators that have been used to calculate the Internet Society Pulse IRI:

1. **Direct indicator:** A direct indicator is a direct measure of an aspect of resilience e.g., percentage of HTTPS adoption, latency, bandwidth, etc. They have a specific unit of measurement, and the raw value can be on different scales depending on what is being measured.
2. **Composite indicator:** A composite indicator provides a score, which itself has been derived from multiple other variables. Examples are the MANRS score, EGDI index, Market Concentration, etc. The scale of a composite indicator is usually between 0 and 100.
3. **Proxy indicator:** A proxy is used where it is difficult to find a specific metric to measure an aspect of resilience. Proxies can be either direct or composite indicators. For example, the IRI uses "Number of IXPs" and "Number of data centers", together to quantify the robustness of the local infrastructure.

Orientation of Indicators

An indicator can either be positive or a negative. In the Internet Society Pulse IRI framework, both positive and negative indicators are used either individually or in combination with other indicators to depict a notion of resilience. An example of a positive indicator is "Number of secure Internet servers" as the higher the number the more secure will be the network. On the contrary, "% of spam infections" is a negative indicator, as the higher the percentage, the less secure the underlying networks are.

Details on Some Indicators

10-km Fiber reach

This dataset is the population within a 10, 25 and 50-km range of operational core transmission network fiber nodes only. The number of nodes included in the map depends entirely on the level of detail in the source material available, which varies from operator to operator. It does not include fiber metro or access (FTTx) networks, because street-level detail of fiber optic cabling and building endpoints are almost never disclosed by network operators. See Broadband Transmission Capacity Indicators Definitions - Indicator 7a Percentage of population within reach of transmission networks) https://www.itu.int/en/ITU-D/Technology/Documents/InteractiveTransmissionMaps/Misc/BroadbandTransmissionCapacityIndicators.pdf

Network Performance

The data about bandwidth and latency is collected from the monthly Ookla Speedtest Global Index[5]. It contains measurements about fixed and mobile network performance around the world. The median download, upload and latency values are calculated by country.

Upstream redundancy

The Upstream redundancy is the average number of upstream providers by active Autonomous Systems (ASes) in the country (weighted by market share). The higher the number of upstream providers per AS, the more resilient the overall ecosystem is. The CAIDA AS-Relationship[6] dataset is used to infer the provider to customer relationship.

Peering Efficiency

The Peering Efficiency score of a country is calculated by taking the sum of all unique and local[7] ASes peering at an IXP in a country and dividing it by the number of allocated and active (seen on the global routing table) ASes in a country. PeeringDB[8] or Packet Clearing House (PCH)[9] provides the number of unique local peers and the RIRs delegated file[10] provides the total number of allocated ASes by country.

$$PE_c = \frac{\sum P_i}{A}$$

Where:

$$PE_c = Peering\ Efficiency\ of\ country\ c$$

$$P_i = Local\ ASes\ peering\ at\ IXP\ i$$

$$A = Number\ of\ allocated\ ASes\ for\ country\ c$$

Market Concentration

The Internet Society Pulse IRI uses the Herfindahl-Hirschman Index (HHI) to calculate the market concentration score. APNIC ASPOP statistics[11] provide market share information by AS and by country.

5 https://www.speedtest.net/global-index
6 https://www.caida.org/catalog/datasets/as-relationships/
7 Both local and foreign ASes (e.g., CDNs, Tier-1 operators) peering at an IXP. For this calculation, only local ASes are considered.
8 https://www.peeringdb.com/
9 https://www.pch.net/ixp/dir
10 https://ftp.ripe.net/pub/stats/ripencc/nro-stats/latest/nro-delegated-stats
11 https://stats.labs.apnic.net/cgi-bin/aspop

The HHI has a range between 0 and 10,000, where 0 means no concentration (a competitive market) and 10,000 means only one ASN is present i.e., with 100% market share.

$$HHI_c = s_1^2 + s_2^2 + s_3^2 + \cdots s_n^2$$

Where:

$$HHI_c = HHI\ index\ of\ country\ c$$
$$s_n = market\ share\ (\%)\ of\ ASN_n\ of\ country\ c$$

AS Hegemony

Network centralization is an important element to measure as it indicates the extent to which the relationships of a given network are concentrated on a single network or group of networks. At a country-level, there are specific network operators providing international access and the more diverse the number of upstream Internet providers, the more resilient the country is in terms of network dependency.

The notion of network dependency can be proxied using AS Hegemony[12] which is a score given to a network to quantify its centrality as observed by BGP monitors. AS hegemony ranges between 0 and 1 and can be interpreted as the average fraction of paths crossing a node. The higher the AS Hegemony score, the higher the dependency on that specific network.

Each network in a country has an AS Hegemony score, based on how central it is for all other eyeball networks. To calculate the inequality in the network dependency distribution at a country-level, we use the GINI coefficient[13] of inequality. In a perfectly equal scenario (GINI = 0), all networks would have the same AS Hegemony score. GINI=1 means perfect inequality.

Examples:

- Country C has three networks AS X (10% coverage), AS Y (50% coverage), AS Z (40% coverage). The GINI coefficient of country C is G(5, 10, 85) = 0.533 (High GINI)
- Country D has three networks AS X (33% coverage), AS Y (33% coverage), AS Z (34% coverage). The GINI coefficient of country D is G(33, 33 , 34) = 0.07 (Low GINI)

[12] https://www.iij-ii.co.jp/en/members/romain/pdf/romain_sigcomm2017.pdf
[13] https://en.wikipedia.org/wiki/Gini_coefficient

List of indicators

Table 1 shows the list of indicators, the unit of measure and the source of the information.

Indicator	Description	Unit	Source
10-km Fiber reach	Percentage of the population within 10 km of a fiber connection point	% of population	ITU
Network Coverage	Mobile Network coverage includes 2G/3G/4G with a composite score provided by the GSMA	Score (0 - 100)	GSMA
Spectrum allocation	Spectrum allocation (composite score)	Score (0 - 100)	GSMA
Number of IXPs	Number of IXPs (Ratio of the number of IXPs and the number of cities in a country with population of > 300,000)	# of IXPs per city	PCH PeeringDB
Data centers	Number of datacenters	# of datacenters per 10 million population	PeeringDB
Mobile/Fixed Latency	Median latency observed to the nearest Ookla server	ms	Ookla
Mobile/Fixed Jitter	Median Jitter observed to the nearest Ookla server	ms	Ookla
Mobile/Fixed Upload speed	Median upload throughput measured to the nearest Ookla server	Mbps	Ookla
Mobile/Fixed Download speed	Median download throughput measured to the nearest Ookla server	Mbps	Ookla
IPv6	IPv6 Enabled end users	% of IPv6 adoption	Akamai, Facebook, Google, APNIC
HTTPS	Pageload using HTTPS	% of pageload on HTTPS	Mozilla

Indicator	Description	Unit	Source
DNSSEC Validation	Users validating DNSSEC	% of users validating DNSSEC	APNIC
DNSSEC Adoption	Is the ccTLD signed?	0 or 1	ICANN
MANRS Readiness	MANRS score (Filtering, Global Coordination, IRR, RPKI)	Score (0 – 100)	MANRS Observatory
Upstream Redundancy	Average number of upstream providers for a country ASN	Score (0 – 100)	CAIDA
Secure Internet Servers	Number of Secure Internet Servers detected on the country's network	# of Secure servers per 1000 population	World Bank
Global Cybersecurity Index	Global Cybersecurity Index (Composite score)	Score (0 – 100)	ITU
DDOS Potential	Potential DDOS threat a country represents	Percentage	Cybergreen
Affordability	Average of affordability for fixed and mobile broadband	% of GNI per capita	ITU/A4AI
Market concentration	Herfindahl-Hirschman Index (HHI) calculates the market concentration based on market share information per network.	Score (0 – 10000)	APNIC
Network Centralization	GINI Coefficient is used to calculate the inequality in the dependency on specific network for upstream connectivity.	Score (0 – 100)	IIJ
Peering efficiency	Ratio of ASes peering at IXPs vs allocated ASes in a country	Percentage	PCH PeeringDB
Domain count	Domains registered by ccTLD	# of domains per ccTLD per 1000 pop.	DomainTools
EGDI	E-Government Development Index	Index (0 – 100)	UN

Table 1. List of indicators

Data Processing

Raw data comes in different forms and shapes and usually comes with several artifacts - some datasets are normally distributed, while some others are skewed. Before running any calculation or aggregation it needs to be imputed for missing data and treated for outliers.

Missing Data

The following techniques have been used to impute missing data:

Indicator	Technique	Details
Affordability	Linear Regression	The affordability value is calculated by using simple regression imputation. Affordability is highly correlated with Internet Penetration (Pearson=-0.80)
Latency, download and upload speed	Substitution	M-Lab speedtest dataset was used as a replacement.
Other indicators	Substitution	Replacement by data from a similar country based on GPD per capita.

Table 2. Data imputation

Re-scaling and Treating Outliers

The scales used by the indicators are also different e.g., latency can range between 0 – 500ms, while domain count for a ccTLD can range between 0 – 2,000,000. It is important to scale the data so that indicators are comparable to one another, and to avoid the issue of the size of the country (i.e., larger countries in terms of population or GDP tend to have more networks, IXPs, datacenters, etc.).

On the other hand, outliers have the tendency to skew the data and can therefore have an impact on the overall score calculation, especially that Internet Society Pulse IRI uses the min-max normalization (see section on Min-Max Normalization below) method to scale the data. If an indicator has a very high or very low value, this will be reflected in the min-max calculation.

The following transformation has been applied to the specific indicators of the framework:

1. Denomination by population size: Number of data centers, Number of domains
2. Denomination by number of cities: Number of IXPs
3. Log transformation[14]: HHI Index, Secure Internet Servers, fixed/mobile upload speed, fixed/mobile download speed and fixed/mobile latency.

After scaling and transforming the above indicators, we run a check on the skewness and kurtosis values on the remaining indicators. For those having a skewness > 2 and kurtosis > 3.5 (general threshold for outlier detection), the IRI makes use of the IQR (Interquartile Range: Q3 -Q1) method to trim down outliers. The following rules are applied:

- Any value greater than Q3 + 1.5*IQR, is replace by Q3 (75th percentile)
- Any value less than Q1 – 1.5*IQR, is replaced by Q1 (25th percentile)

Min-Max Normalization

The next step, after cleaning and transforming the data, is normalization. Normalization is important because indicators are collected using different unit of measurements (percentage, ms, Mbps, count, etc.). It is therefore important to rebase them to a common unit such as into 0 to 100 scale, where 100 usually refers to the strongest and 0 to the weakest value.

The method chosen was the *min-max* normalization, it is a common technique used by multiple known indices[15][16] and as opposed to other techniques such as ranking and categorical scales, *min-max* keeps the interval between the countries consistent.

Below are the formula Internet Society Pulse IRI uses to calculate the value of an indicator depending on whether it is positive or negative:

Positive indicator	Negative indicator
$\boldsymbol{I_{k,c} = \frac{x_{k,c} - Min(x_k)}{Max(x_k) - Min(x_k)}}$	$I_{k,c} = 1 - \frac{x_{k,c} - Min(x_k)}{Min(x_k) - Min(x_k)}$

[14] A Logarithmic transformation is useful to treat skewed datasets and to discard extreme values. Not only it scales the data, but it has the advantage of handling outliers in the dataset. Log transformation preserves the differences between the values.

[15] EIU Internet Inclusive Index - https://theinclusiveinternet.eiu.com/

[16] GSMA Mobile Connectivity Index - https://www.mobileconnectivityindex.com/

where "x" refers to the raw value of the indicator "k" of country "c" and "I" refers to the normalized value. $Max/Min(x_k)$ refers to the min/max of indicator "k" for all countries.

Positive indicators contribute towards increasing an index, negative indicators contribute to decrease the score, which is why we take the delta $(1 - I_{k,c})$.

Finally, we chose not to use the *z-score* standardization[17] technique as not all the indicators followed a normal distribution.

Weighting and Aggregation

Assigning Weights

There are two main ways to aggregate the normalized indicators into a final score using:

1. An ad-hoc weighting scheme.
2. Statistical (optimization) techniques.

The Internet Society Pulse IRI uses a weighting scheme as it is the simplest technique between the two and relies on input that the Internet Society gathered through survey and discussions with subject matter experts.

During the weighting process, the importance of the indicator was also considered using a "lifecycle" approach. For example, for the Infrastructure pillar, the following weights were assigned to the underlying dimensions: Cable Ecosystem (40%), Mobile Connectivity (30%) and Enabling infrastructure (30%). Higher importance was given to *Cable Ecosystem* as it is a prerequisite for a functional Internet.

In the Internet Society Pulse IRI framework, the indicators are grouped into different dimensions, and the dimensions into pillars, which on their own provide quantitative measures of a specific aspect of Internet resilience. Below is a table showing the indicators, dimensions and pillars and their associated weights, used for the calculation of the Internet Society Pulse IRI in 2023.

[17] This technique standardizes around the mean value and ranges between 0 and 1.

The weights will be revisited on an annual basis.

Pillar	Weight (%)	Dimension	Weight (%)	Indicator	Weight (%)
Infrastructure	25	Cable Infrastructure	40	10-km Fiber reach	100
		Mobile connectivity	30	Network Coverage	70
				Spectrum allocation	30
		Enabling infrastructure	30	Number of IXPs	50
				Datacenters	50
Performance	25	Fixed networks	40	Latency	20
				Upload	30
				Download	30
				Jitter	20
		Mobile networks	60	Latency	20
				Upload	30
				Download	30
				Jitter	20
Enabling technologies and security	25	Enabling technologies	20	IPv6	30
				HTTPS	70
		DNS ecosystem	30	DNSSEC Validation	50
				DNSSEC Adoption	50
		Routing hygiene	30	MANRS Readiness	50
				Upstream Redundancy	50
		Security threat	20	Secure Internet Servers	30
				Global Cybersecurity Index	40

Pillar	Weight (%)	Dimension	Weight (%)	Indicator	Weight (%)
				DDOS Potential	30
Local ecosystem & Market readiness	25	Market structure	50	Affordability	40
				Market concentration	30
				AS Hegemony	30
		Traffic Localization	50	Peering efficiency	40
				Domain count	30
				EGDI	30

Table 3. Indicators, dimensions and pillars and associated weights

Aggregation

The Internet Society Pulse IRI uses a weighted sum formula at each level (indicator, dimension, and pillar) to aggregate the data into a composite score. The following formula was used:

$$IRI_c = \sum_{i}^{n} (w_i * P_{i,c})$$

Where:

$$P_{i,c} = \sum_{i}^{n} (w_i * D_{i,c})$$

And where:

$$D_{i,c} = \sum_{i}^{n} (w_i * I_{i,c})$$

In simple terms, the final index IRI_c of country "c" is the sum of the weighted pillars "P_i". A pillar is the weighted sum of the underlying dimensions "D_i" and a dimension is the weighted sum of the indicators "I_i" all of country "c".

Feedback

For any questions, comments, and feedback on the Internet Society Pulse IRI, please contact the Internet Society Pulse team (pulse@isoc.org).

Acknowledgements

The Internet Society would like to thank the following contributors for their valuable input in the conception of the Internet Society Pulse Internet Resilience Index (IRI). Amreesh Phokeer (Internet Society), Kevin Chege (Internet Society), Assane Gueye (Carnegie Mellon University-Africa), Josiah Chavula (University of Cape Town), and Ahmed Elmokashfi (Simula Research Lab).