

Technical Report / Teaching Material
TUHHCNTR002

Performance Metrics in Communication Networks Research

Institute of Communication Networks (ComNets), TUHH

Draft Version 0.1¹, May 6, 2026

1 Introduction

Performance metrics are the most common way to numerically evaluate the performance of a given system. To ensure that research efforts from different groups and on different topics can be compared to each other, a set of common and widely used performance metrics has been established by several standardisation bodies like Internet Engineering Task Force (IETF). In this document, a list of selected performance metrics, their definition and instructions on how they are computed is given, as they are used by the Institute of Communication Networks at the TUHH.

1.1 Units used in Communication Networks

Table 1 highlights a few examples of how to write units as defined by the International Systems of Units (SI). Binary data is measured in bit and prefixes follow the metric system. More details can be found in [1].

Note: 1 kbit/s = 1000 bit/s — 1 Mbit/s = 1 000 000 bit/s

2 Elementary Performance Metrics

The most important performance metrics in communication networks are Throughput, Delay and Packet Delivery Ratio as given in the subsequent Table 2. They are defined in detail in subsections 2.1 to 2.3.

¹DOI: <https://doi.org/10.15480/882.17078>

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). You are free to share and adapt the work, provided you give appropriate credit.

Name	Symbol
100 bit	100 bit
100 byte	100 B
100 kilobyte	100 kB
100 megabyte	100 MB
100 gigabyte	100 GB
100 kilobit per second	100 kbit/s
100 megabit per second	100 Mbit/s
100 gigabit per second	100 Gbit/s

Table 1: Use of SI units

Metric	General Description	Unit
Throughput	amount of data packets transmitted in a given amount of time	bit/s
Delay (Latency)	time for data packet to travel between two reference points e.g. from exiting Application Layer at source until entering Application Layer at destination	s
Packet Delivery Ratio	how many data packets were delivered to an intended recipient	%

Table 2: Summary of elementary performance metrics

There are numerous ways to define these performance metrics in literature. How these metrics are defined as in [2] is explained in the following sub-sections. The Appendix shows the specific definitions proposed w.r.t. specific protocols by different standardization bodies.

2.1 Throughput

Throughput refers to the amount of data packets successfully transferred from source to destination within a given period of time T . It is usually measured in bits per second. If the destination receives N number of data packets each of size k bit, the throughput measured at the destination can be expressed as:

$$\gamma = \frac{Nk}{T} \text{ [bit/s]} \quad (1)$$

The throughput at the application layer is also referred as goodput.

2.2 Delay

The total delay is composed of different parts:

$$\delta = \delta_{proc} + \delta_{queue} + \delta_{trans} + \delta_{prop} \text{ [s]} \quad (2)$$

- δ_{proc} : processing delay: the time a node takes to process packet headers typically a few micro-seconds or less
- δ_{queue} : queuing delay: the time a packet has to wait in a buffer of the originating node or at the intermediate node until it has an opportunity to be transmitted. It is variable and depends e.g. on network congestion.

- δ_{trans} : transmission delay: the time required to transmit a complete packet or message on the transmission medium. The transmission delay is calculated as:

$$\delta_{trans} = \frac{\text{Packet size}}{\text{Data rate}} [\text{s}]$$

- δ_{prop} : propagation delay: the time it takes for the signal to physically travel from the transmitter to the receiver, typically a few microseconds to hundreds of milliseconds, depending on the distance and the transmission medium. The propagation delay is calculated as:

$$\delta_{prop} = \frac{\text{Transmission distance}}{\text{Signal velocity}} [\text{s}]$$

It is tedious to measure the above mentioned delay components independently in practice. Therefore, the total delay is measured in simulations or experiments by computing the time difference between the packet reception and transmission time. The total delay is also referred as end-to-end delay, one time delay, delivery delay, etc. The end-to-end delay is often measured via the Round-Trip-Time (RTT), where the delay is measured from the time difference between transmission and reception of feedback from the destination at the source e.g. acknowledgments.

The delay provides a metric of how fast a message can be delivered to an intended recipient (or a group of recipients) considering a per-node or a per-network scope. The timeliness of data can be evaluated as *Average delay to receive certain data per node* and *Average time to propagate data through the network*.

The *Delay* δ_i at the i^{th} node of a network is computed according to Equation (3).

$$\delta_i = \frac{1}{M_{rx,i}} \sum_{j=1}^{M_{rx,i}} (\tau_{rx,i,j} - \tau_{tx,i,j}) [\text{s}] \quad (3)$$

where $M_{rx,i}$ is the total number of data packets (i.e., messages or data) received by the i^{th} node in the network. The transmission time $\tau_{tx,i,j}$ is the time when the source generates the j^{th} data packet. The reception time $\tau_{rx,i,j}$ refers to the time when the j^{th} data packet is received by the node i .

The network-wide *Mean Delay*, $E[\delta]$ is consequentially computed as shown in Equation (4) where N is the number of nodes in the network.

$$E[\delta] = \frac{1}{N} \sum_{i=1}^N \delta_i [\text{s}] \quad (4)$$

In the case where a data packet is destined to all recipients (broadcast scenarios), the delay per node is computed according to Equation (5).

$$\delta_i = \frac{1}{M} \sum_{j=1}^M (\tau_{rx,i,j} - \tau_{tx,j}) [\text{s}] \quad (5)$$

where M is the total number of data packets created in this network, $\tau_{tx,j}$ is the time the j^{th} data packet was generated and $\tau_{rx,i,j}$ is the time node i received it. The Equation (5) is valid under the assumption that each node receives all the data packets created in the network.

In addition, it is often helpful to know the distribution of the different delays as their significance can vary between scenarios and protocols. Therefore, it is beneficial to plot the Cumulative Distribution Function (CDF) of the delays of individual nodes and all nodes in the network.

2.3 Packet Delivery Ratio

The *Packet Delivery Ratio* or *Packet Reception Ratio* provides a metric of how many data packets were delivered to an intended recipient. The *Packet Loss Ratio* is the contrary metric, which states how many data packets were lost on the way to an intended recipient. One reason for the loss of a data packet in the network is due to the removal from buffers of the intermediate routers due to network congestion or reaching its expiration time, i.e., Time-To-Live (TTL). Another reason are transmission errors, which are common in wireless communication.

The Packet Delivery Ratio of the i^{th} node is computed according to Equation (6).

$$\eta_i = \frac{M_{rx,i}}{M_{tx,i}} \quad (6)$$

where $M_{tx,i}$ is the total number of data packets created by all the nodes in the network destined to node i and $M_{rx,i}$ is the number of packets received by the intended recipient node i .

Equation (6) is also valid in the case where a data packet is destined to multiple recipients (multicast or broadcast scenarios). Here, $M_{tx,i}$ refers to the total number of data packets created in the network.

The network-wide *Mean Packet Delivery Ratio*, $E[\eta]$, is computed in Equation (7).

$$E[\eta] = \frac{1}{N} \sum_{i=1}^N \eta_i \quad (7)$$

where N is the number of nodes in the network and η_i refers to the Packet Delivery Ratio of the i^{th} node.

3 Protocol Specific Metrics

3.1 Overhead

The term overhead in general refers to the additional activities needed to achieve an intended objective. The overhead is a measure of how much of these additional activities are required to deliver data packets to the intended recipients. The overhead can be shown as it is or as a ratio.

3.1.1 Control Packet Overhead

The *Control Packet Overhead Ratio* of the i^{th} node is computed according to Equation (8).

$$\theta_i = \frac{M_{tx,i}}{M_{tx,i} + C_{tx,i}} \quad (8)$$

where $M_{tx,i}$ is the total number of data packets (or bytes) transmitted by the node i and $C_{tx,i}$ is the number of control packets (or bytes) transmitted by the node i . The control packets should be defined w.r.t. your scenario. For example, if you want to show the routing overhead, all the additional packets generated by the routing protocol are considered as control packets.

The network-wide *Mean Control Packet Overhead Ratio*, $E[\theta]$, is computed in Equation (9).

$$E[\theta] = \frac{1}{N} \sum_{i=1}^N \theta_i \quad (9)$$

where N is the number of nodes in the network and θ_i refers to the *Control Packet Overhead Ratio* of the i^{th} node.

3.1.2 Protocol Overhead

The *Protocol Overhead Ratio* for the j^{th} data packet indicates how much header information is added to transmit the payload as shown in Equation (10).

$$\theta_j = \frac{P_j}{P_j + H_j} \quad (10)$$

where P_j is the size of the payload (in bytes) of the j^{th} data packet and H_j is the size of the headers (in bytes) added when transmitting the j^{th} data packet from the source to the destination.

4 Application Specific Metrics

4.1 Age of Information

In distributed real-time and sensor applications, such as vehicular networking and networked control systems, not only the node delay is of relevance. Instead, the freshness of sensor or control knowledge on the receiver side adds additional evidence of system performance, since it reduces delays, consecutive transmission losses, and transmission intervals. This section details one of the ways to express the impact of delayed and outdated data on a system by a performance metric called age of information [3].

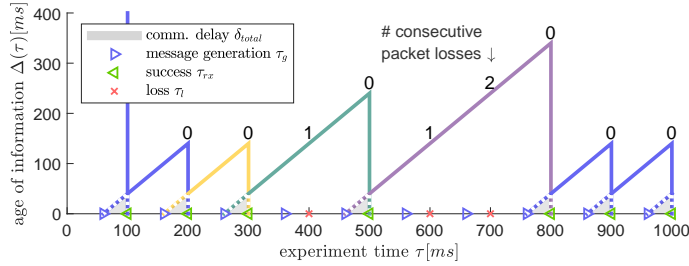


Figure 1: Age of information over time in an example with 1 intermediate loss for Δ_j (green) and 2 consecutive losses for Δ_{j+1} (purple), with $t_{sent} = 100$ ms and $\delta_{total} = 40$ ms. The age of information is infinite before the first successful transmission at $\tau = 100$ ms.

The age of information over time $\Delta(\tau)$ is defined as follows. Let δ_j be the total delay of the communication from source to destination for a *successful* transmission at time $\tau_{rx,j}$. The total delay of the j^{th} data packet, $\delta_j = \delta_{proc,j} + \delta_{queue,j} + \delta_{trans,j} + \delta_{prop,j}$, depends on the processing, queuing, transmission and propagation delays (see section 2.2). For any time instance τ and for the last successful transmission at time $\tau_{rx,j}$, the age of information $\Delta(\tau)$ is defined in Equation (11).

$$\Delta(\tau) = \tau - \tau_{rx,j} + \delta_{total,j} \text{ [ms]} \quad (11)$$

Between the time of $(j)^{\text{th}}$ and $(j+1)^{\text{th}}$ packet reception, $\tau_{rx,j}$ and $\tau_{rx,j+1}$ respectively, $\Delta(\tau)$ grows linearly up to a maximum of Δ_j . Afterwards, the age of information is decreased by a step function and takes the new value $\delta_{total,j+1}$ (see Figure 1). Since it continues to grow during unsuccessful transmission attempts k and grows higher for longer gaps between message generation τ_{sent} , the age of information metric includes both the impact of packet losses and transmission intervals (see Figure 2).

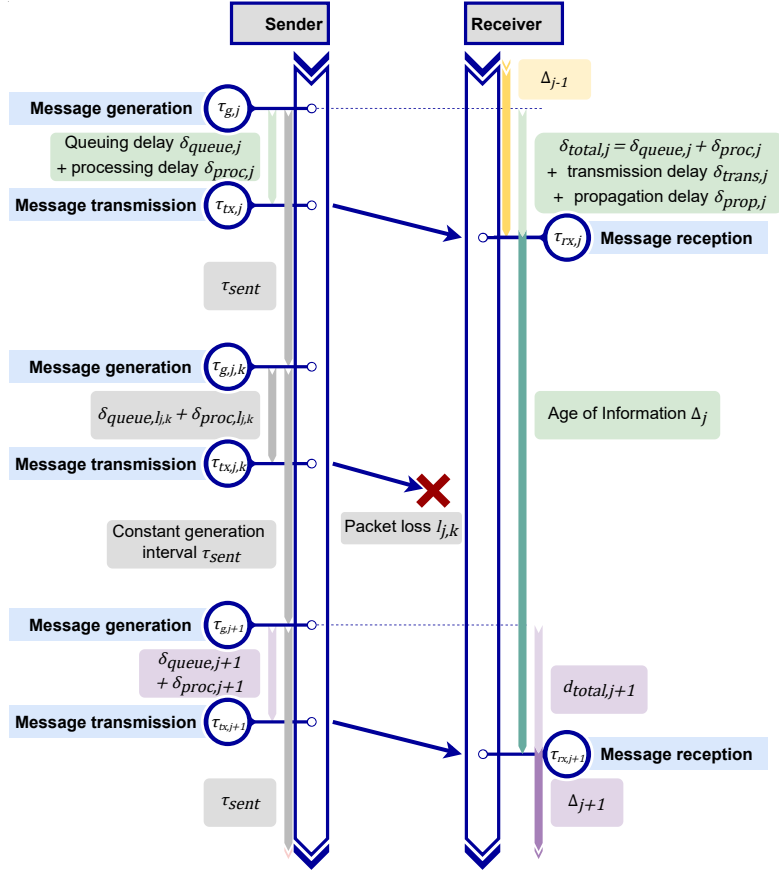


Figure 2: Message sequence chart of the age of information of a transmission sequence with an intermediate loss $l_{j,k} = 1$ for Δ_j (green).

4.2 Another Dimension to Performance Metrics

The performance metrics given above show the technical characteristics of a network. A special set of performance metrics and requirements related to industrial applications, where real time control is achieved using communication networks can be found in [4].

- Reliability: ability of an item to perform a required function under given conditions for a given time interval
- Dependability: ability to perform as and when required
- Availability: ability to be in a state to perform as required
- Maintainability: ability to be retained in, or be restored to a state to perform as required
- Recoverability: ability to recover from a failure, without corrective maintenance

5 Summary

In this report basic performance metrics used in communication networks are discussed. While it can be sufficient to simply provide the aforementioned metrics in a cohesive way for your work, often more sophisticated metrics have to be established. In general, using different metrics (and by extension the ones presented here) always comes with the need for a clear and interpretation free definition. This definition has to be provided by the author and ensures that readers can fully grasp the meaning of the results. Furthermore, in most cases the value of these metrics is not constant but a random variable, so it has to be described by its statistical properties e.g. mean, median, variance, standard deviation, quantiles or distribution function (PDF or CDF).

Once you start your research, you need to be clear what kind of performance metrics are useful for evaluating your specific scenarios. The set of metrics that measure your main research objectives are called Key Performance Indicators (KPIs).

6 Contributors

The contributors to this report are Leonard Fisser, Konrad Fuger, Koojana Kuladinithi, Aliyu Makama, Daniel Plöger, Daniel Stolpmann and Andreas Timm-Giel.

A Appendix - Performance Metrics defined by the other standardisation bodies

A.1 Performance metrics on Internet Protocol (IP) layer by the IETF

- **RFC 2679 - A One-way Delay Metric for IP Performance Metrics (IPPM):** Time passed between “putting” the first bit of the reference packet on the medium at the source node until the last bit was received (“collected” from the medium) by the destination node. Undefined if packet did not arrive (was lost). Requires synchronized clocks.
- **RFC 2681 - A Round-trip Delay Metric for IPPM:** Like RFC 2679 but the destination immediately responds with a packet after receiving the reference packet from the source. Practically more suitable since all measurements can be performed at the source.
- **RFC 3393 - IP Packet Delay Variation Metric for IPPM:** Difference in one-way-delay of two selected reference packets. One-way delay defined according to RFC 2679.
- **RFC 5136 Network Capacity:**

”IP [...] Link Capacity: We define the IP-layer link capacity, $C(L, T, I)$, to be the maximum number of IP-layer bits that can be transmitted from the source S and correctly received by the destination D over the link L during the interval $[T, T + I]$, divided by I .”

”IP [...] Path Capacity: Using our definition for IP-layer link capacity, we can then extend this notion to an entire path, such that the IP-layer path capacity simply becomes that of the link with the smallest capacity along that path. $C(P, T, I) = \min\{1..n\}\{C(L_n, T, I)\}$, L_n is the n 'th link along path P .”

RFC 5136 also defines “link usage”, “link utilization”, “available link capacity”, and “available link usage”

A.2 Performance metrics by the ITU

- **ITU-T Y.1540 - IP packet transfer and availability performance parameters**

”*IP packet transfer reference events*”:

- Packet enters host
- Packet exits host

”*IP packet transfer outcomes*”:

- Successful packet transfer
- Errored packet (failed consistency check at destination)
- Lost packet (never arrived at destination)
- Spurious packet (unexpected packet received “out of nowhere”)
- In-order / reordered (arrived in different order than transmitted)
- ...

”*IP packet transfer performance parameters*”:

- Packet transfer delay (for a single packet)
- Mean packet transfer delay (average over multiple packets)
- Minimum packet transfer delay (lowest experienced delay over multiple packets)
- “End-to-end 2-point [...] packet delay variation” (delay difference relative to reference delay; common reference delay: Minimum packet transfer delay)
- Packet error ratio, packet loss ratio, spurious packet rate, ...
- Capacity: Defined as in RFC 5136 + definition of statistics: average, standard deviation, quantile

ITU G.1000-G.1999 - Multimedia Quality of Service and performance – Generic and user-related aspects (performance as seen by the end-user / application)

- **G.1021** ”Buffer models for development of client performance metrics” (describe receive buffers to compensate delay variation and recovery from out of order packet reception; extends Y.1540 to be applicable for end-user / application performance metrics)
- **G.1030** ”Estimating end-to-end performance in IP networks for data applications” (general considerations for data traffic and example for web-browsing performance metrics)

References

- [1] *Data Rate Units*. https://en.wikipedia.org/wiki/Data-rate_units (Accessed on: 22.09.2020).
- [2] Jens Dede, Anna Förster, Enrique Hernández-Orallo, et al. “Simulating Opportunistic Networks: Survey and Future Directions”. In: *Networking and Internet Architecture - IEEE Communications Surveys and Tutorials* (December 2017), pp. 1–28.
- [3] D. Plöger, M. Segata, R. L. Cigno, et al. “Markov-modulated Models to Estimate the Age of Information in Cooperative Driving”. In: *2019 IEEE Vehicular Networking Conference (VNC)*. 2019, pp. 1–4.
- [4] *Aspects of Dependability Assessments by Industrial Radio Group*. https://industrial-radio-lab.eu/wp-content/uploads/pubs/ZDKI-FG1_Zuverl%C3%A4ssigkeit_EN_210917_v2.pdf (Accessed on: 22.09.2020).