Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Institute of Computer Science
Chair of Communication Networks
Prof. Dr. Tobias Hoßfeld

Chapter 1

Introduction in Performance Modeling

Performance Evaluation of the Internet of Things (IoT)

Module Course: Performance Evaluation of Distributed Systems

Prof. Tobias Hoßfeld, Summer Semester 2022

Disclaimer and Copyright Notice

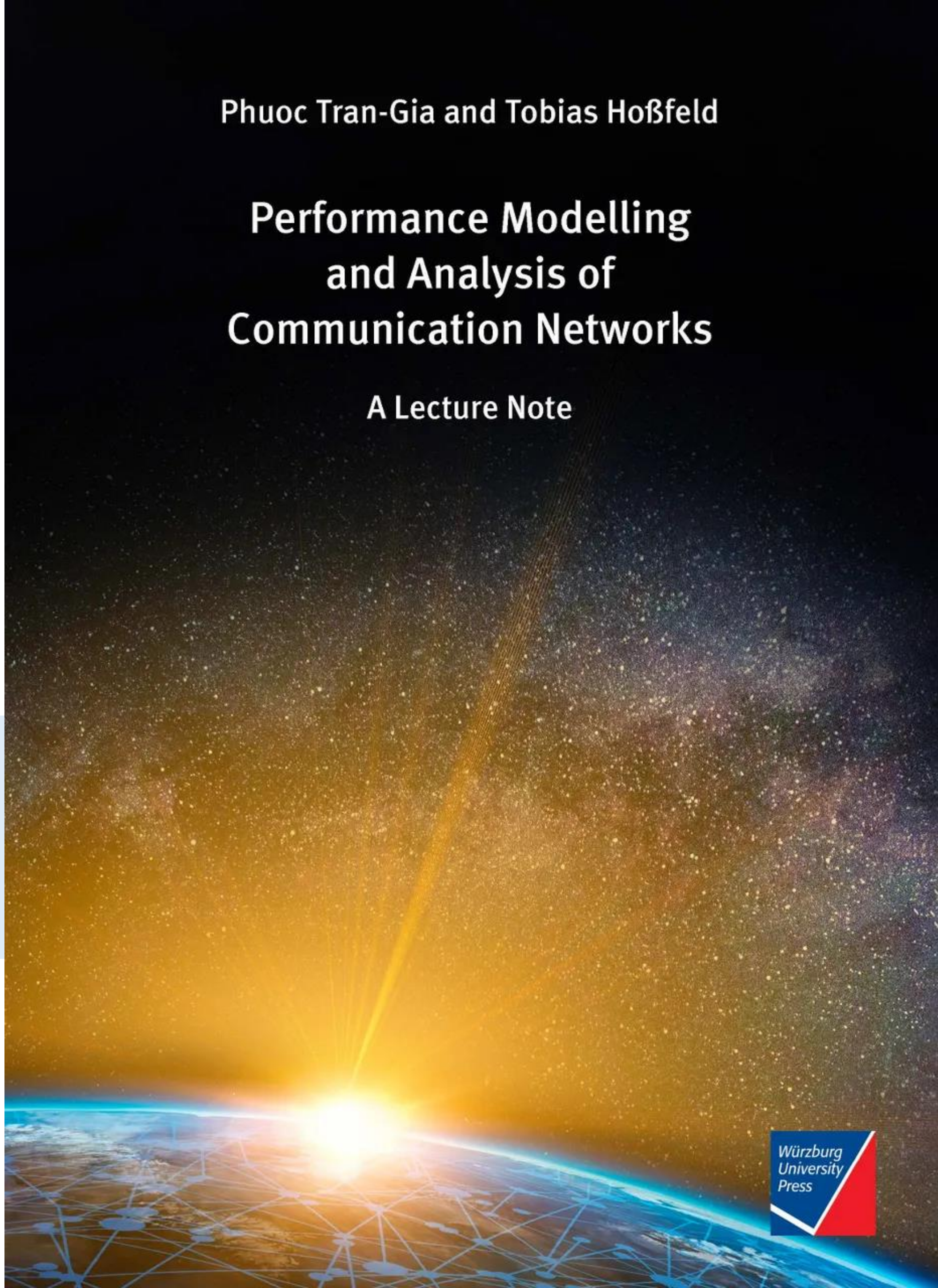
Lecture slides, figures, and scripts are based on the open access text book “Performance Modeling and Analysis of Communication Networks”. The book and scripts are licensed under the Creative Commons License Attribution-ShareAlike 4.0 International ([CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/)). If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

The book must be cited and the disclaimer attached when using lectures slides or scripts.

Tran-Gia, P. & Hossfeld, T. (2021).
Performance Modeling and Analysis of Communication Networks - A Lecture Note. Würzburg University Press.
https://doi.org/10.25972/WUP-978-3-95826-153-2

Website to download book, exercises, slides and scripts:
https://modeling.systems/

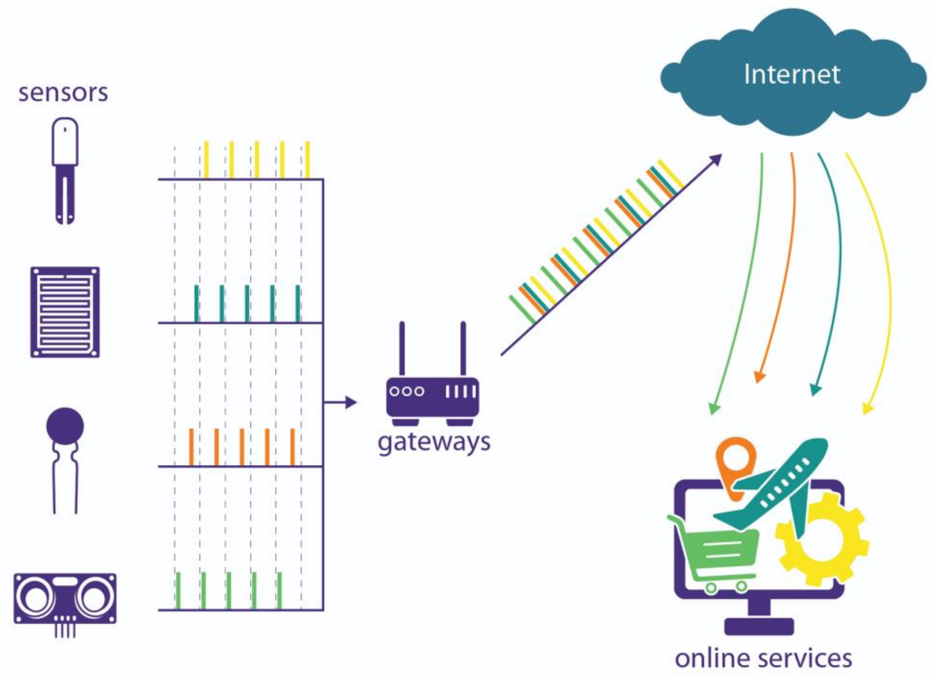
Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG



Chapter 1

Example: Internet of Things (IoT)

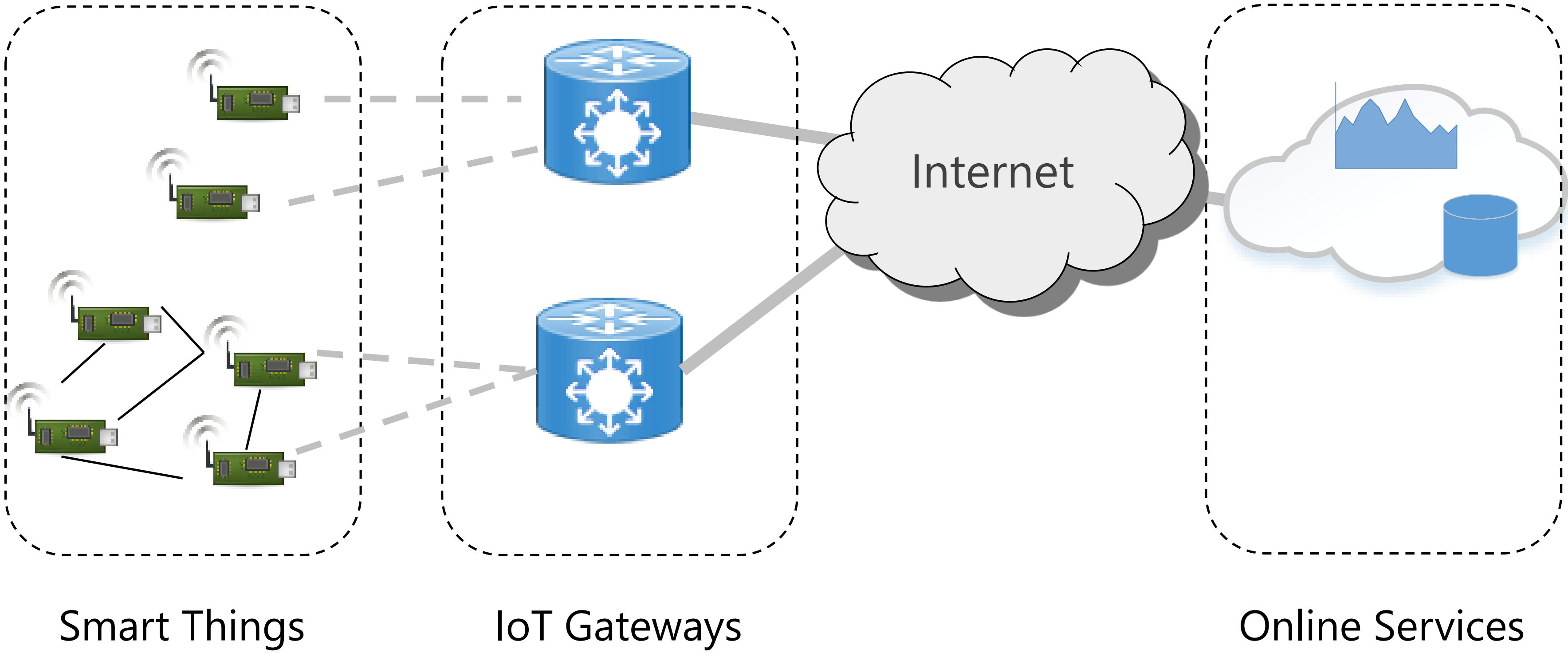
- Some relevant questions
  - How can an IoT system with millions of sensors be modeled? What is the scalability of such a system?
  - How many gateways are needed in a smart city to ensure good quality of service?
  - How can server capacity be dimensioned in a cloud? How many servers or VMs are required?
  - How well does channel access mechanism work for technologies such as nB-IoT or LoRaWAN? What is the probability of a successful data transfer? What is the energy efficiency of the channel access?



- Models from queuing theory applied to IoT systems

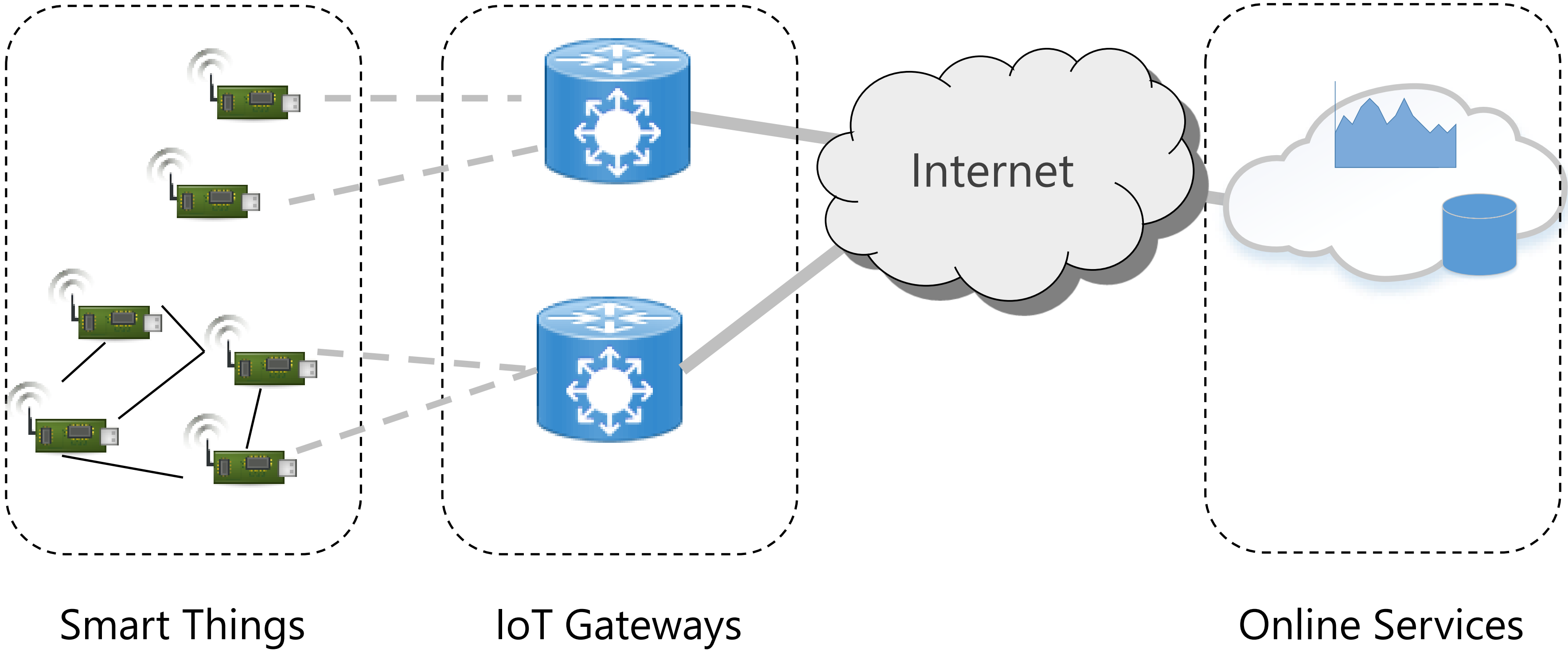
IoT System

- Things networked using Internet technologies
  - may be resource constrained (energy consumption, RAM, ROM)
  - may have to sent a couple of bytes only



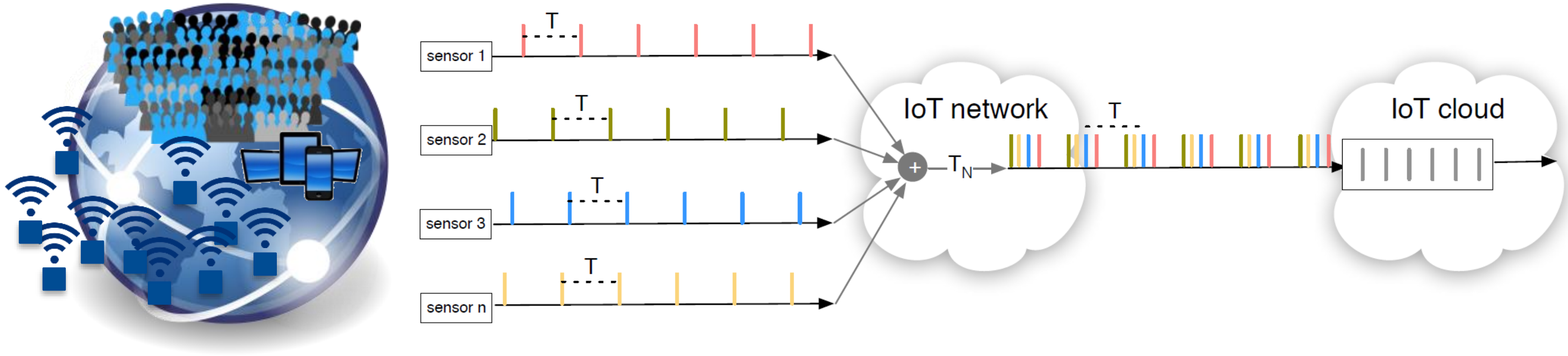
IoT Gateway

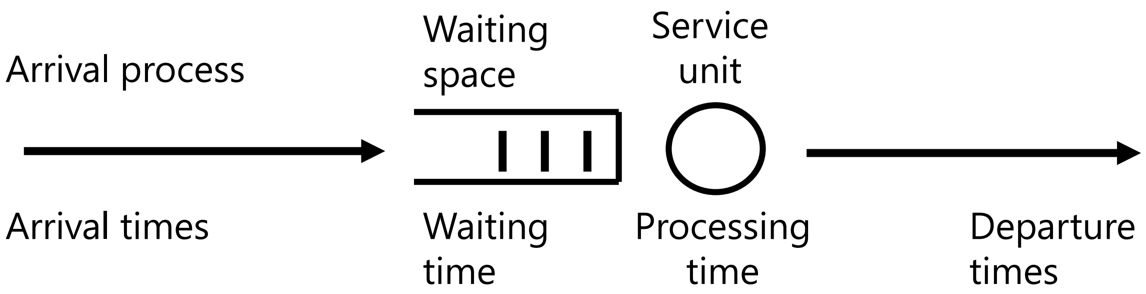
- Bridges the communication gap between IoT devices and the cloud, e.g. translating protocols, adding security features, etc.
- May offer local processing, filtering, caching, storage solutions
- May control smart things
- Advantages: lowering costs, mitigate risks, operation and management

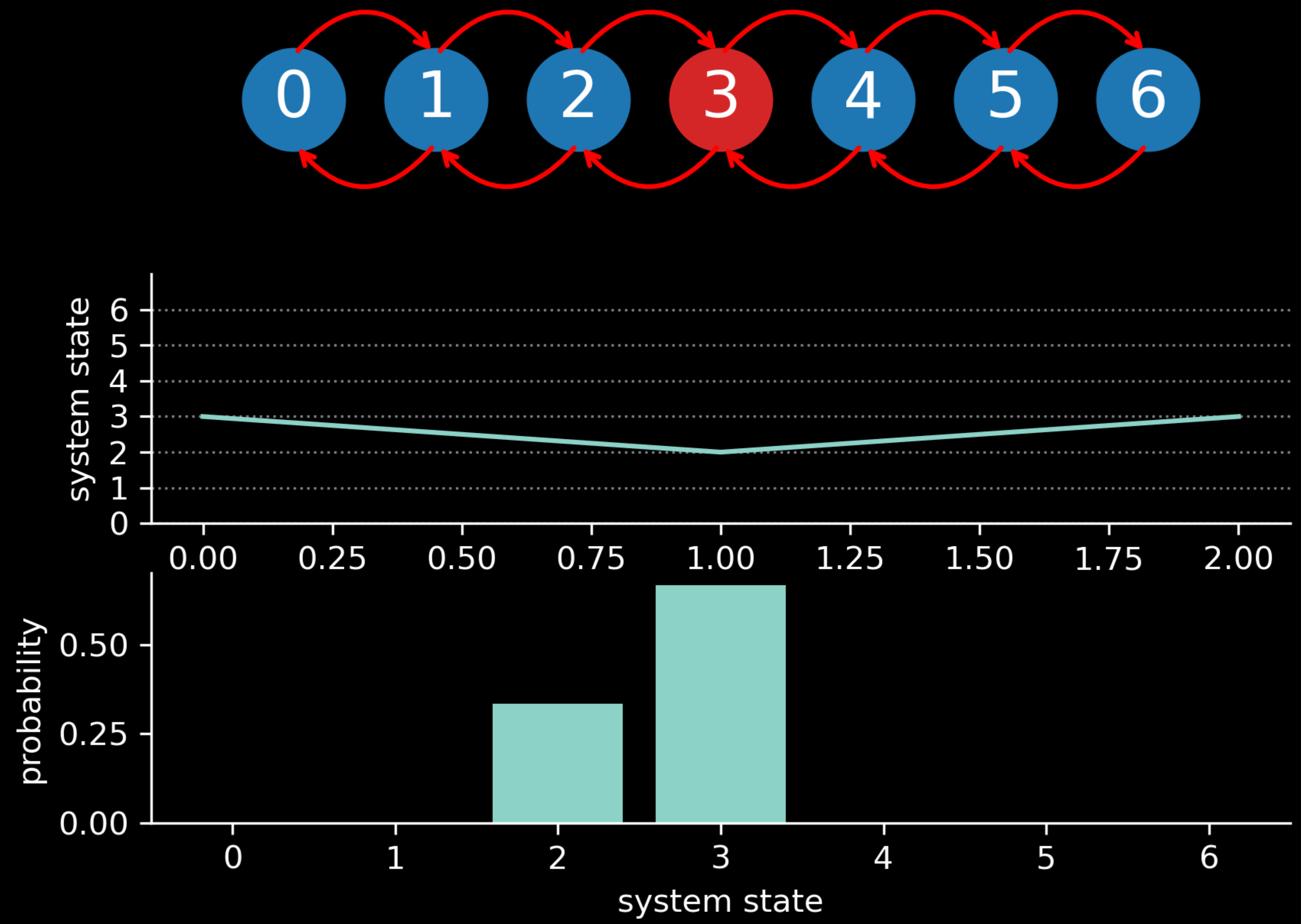


Example: IoT Cloud

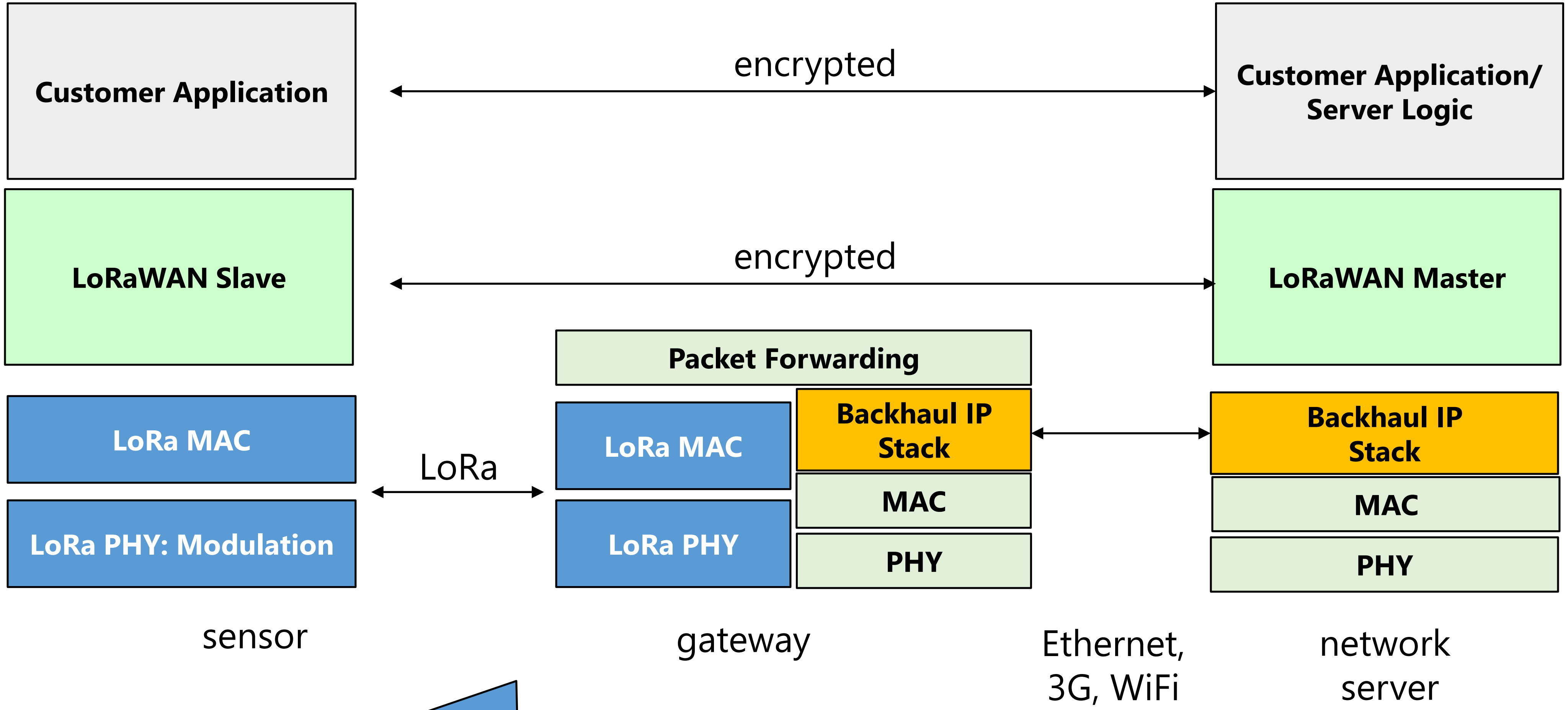
- How to dimension an IoT cloud load balancer? How much traffic from IoT sensors can be handled such that average waiting time is below a treshold? What is the required processing power?






0
1
2
3
4
5
6
system state
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
probability
system state

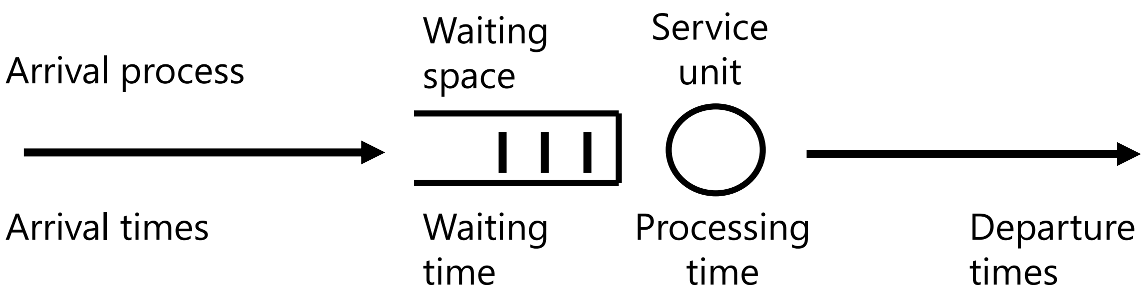
Example: LoRaWAN



What is the performance and energy efficiency of LoRaWAN‘s channel access?

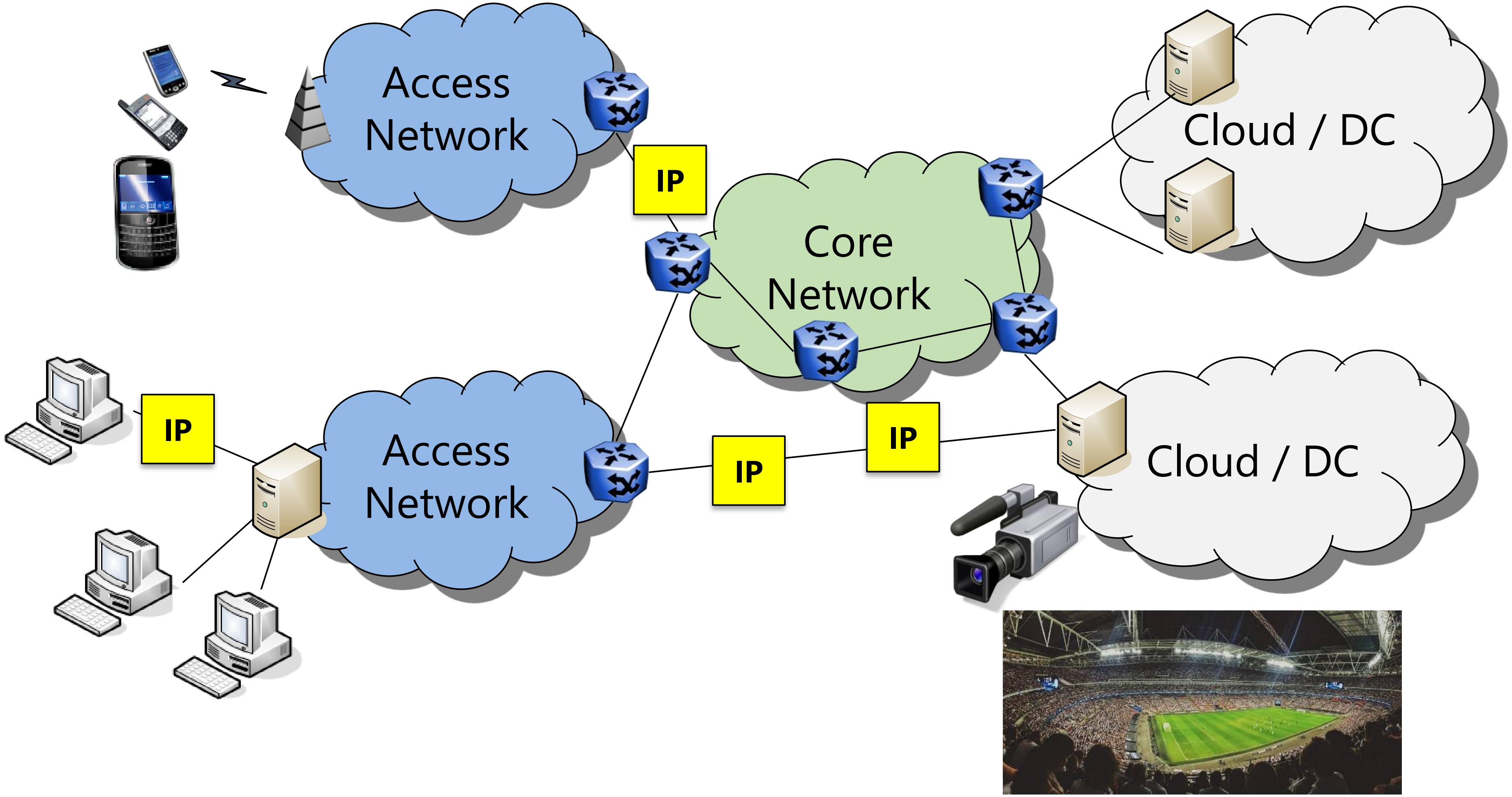
Example: LoRaWAN (f.)

- What are the system components?



Example: Video Streaming

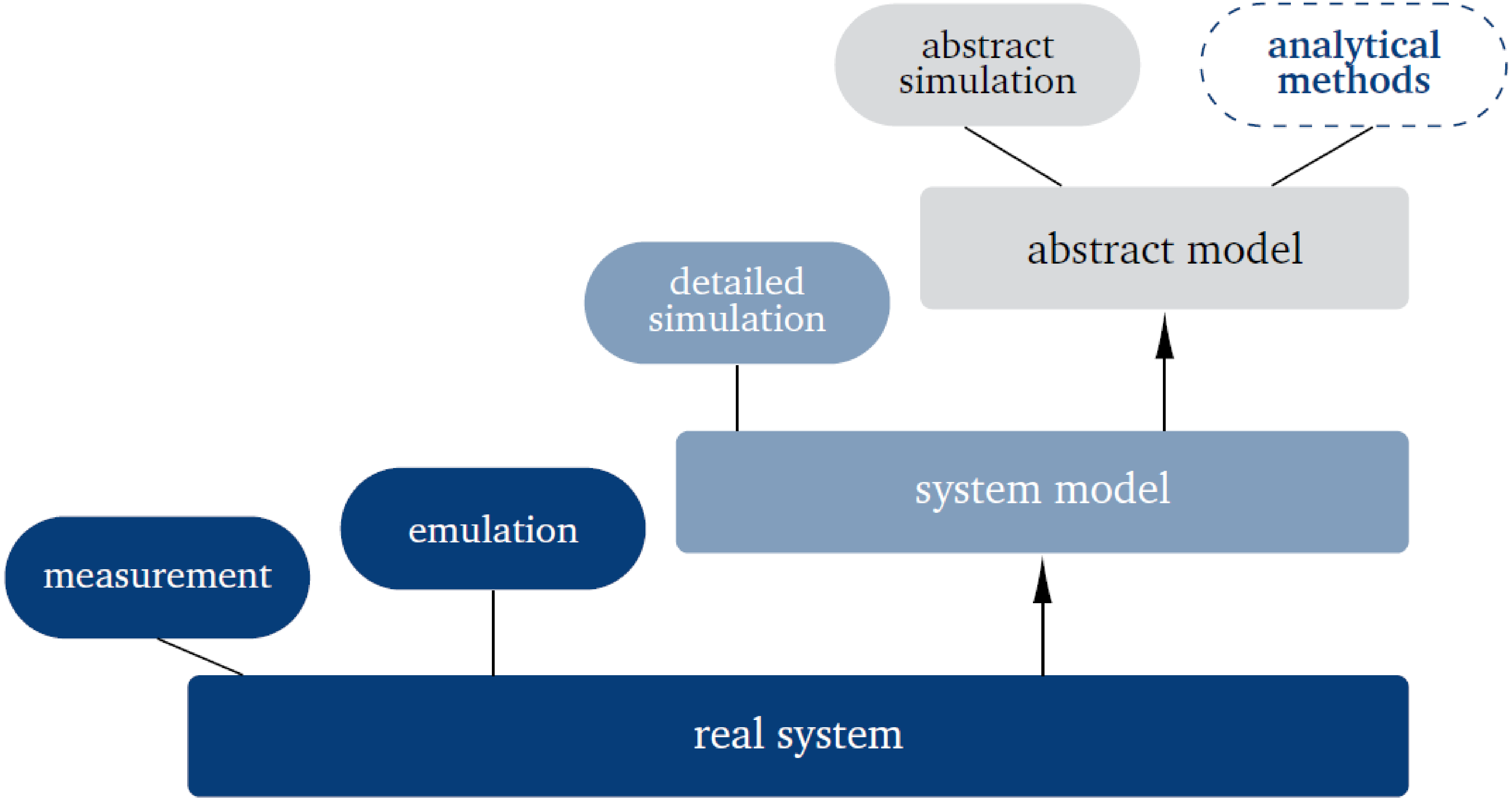
- What is the Quality of Experience of HTTP-based video streaming?



Modeling Concepts

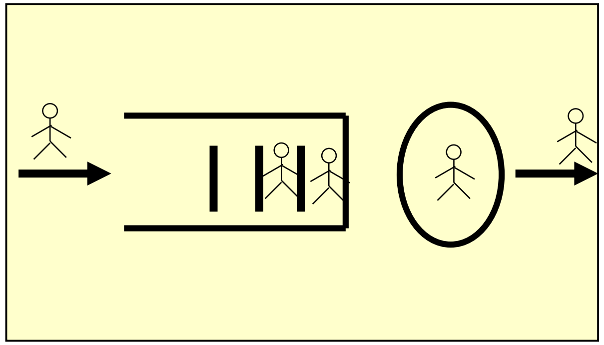
Level of abstraction

Abstraction Layering of Performance Analysis Techniques



Modeling

Real-world problems & questions



real system → abstraction / specification → (formal) model

(formal) model → **Performance evaluation** → solution within the formal model

solution within the formal model → interpretation / transformation → goal: real-world solution & answer

goal: real-world solution & answer → improvement → real system

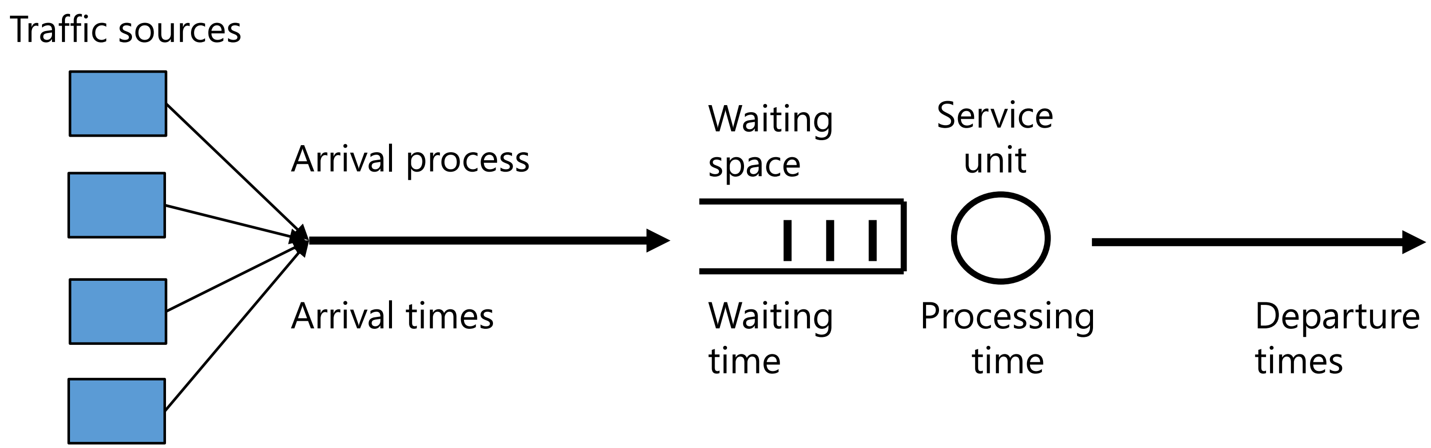
Kendall's Notation

Single-stage queueing systems

Single-Stage Queueing Models

- **Single-stage:** Jobs or customers arrive at the queue, get processed, and leave the system.

- Queueing model requires the definition of the following components:
  - **Traffic sources** and associated random **arrival processes** describing the time between arrivals of jobs in the system.
  - **Service units** (or servers) and associated random variables for the duration of **service processes**.
  - **Queues** and associated waiting room capacity and **operating disciplines**.



Kendall's Notation

- Single-stage queueing models can be described using Kendall's notation

$$GI^{[X]}/GI^{[Y]}/n\text{-}S$$

- S: number of waiting places
  - $S=0$: loss system
  - $0<S<\infty$: delay-loss system
  - $S=\infty$: delay system
- n: number of servers
- $[Y]$: indication of bulk queue *[optional]*
- GI (second): type of service process
- $[X]$: indication of batch arrivals *[optional]*
- GI (first): type of arrival process

Extended Kendall's Notation

A/B/n/S/m/D

A – arrival process: interarrival times (IAT); e.g. M for Poisson process

B – service process: processing times; e.g. D for deterministic service times

n – number of servers or service units; e.g. $n = 1$ for single server system

S – capacity of waiting queue; e.g. loss system $S = 0$; e.g. delay system $S = \infty$

m – size of customer population; default $m = \infty$

D – service discipline, e.g. FIFO (First In – First Out), PS (Processor Sharing)

Remark: Extended Kendall's Notation

- **Compact notation** A/B/n-S
  - specifies A, B, n, S,
  - assumes $m = \infty$ and D = FIFO.
  - Extended notation is then A/B/n/S/∞/FIFO.

- **Extended notation**: Some textbooks use the total capacity of the system $S + n$ instead of the capacity S of the waiting queue only: A/B/n/**S**/m/D
- **Clarification**: We refer to the capacity of the waiting queue and use A/B/n/S/m/D or the short notation A/B/n-S.

- Example:
  - The system with total capacity S+n is written as A/B/n/n+S/m/D.
  - The loss system A/B/n implies A/B/n-0 or A/B/n/0/∞/FIFO.
  - The delay system A/B/n implies A/B/n-∞ or A/B/n/∞/∞/FIFO.

Notation for Arrival and Service Processes

- GI – General independent
  - Arrival processes with renewal properties
  - Arrival or service process can be described with a random variable
  - Realizations of these random variables are statistically independent of one another

- D – Deterministic

- M – Markov
  - Corresponding random variable follows an exponential distribution
  - A M arrival process is a Poisson Process

- E_k – Erlang-k distribution

- H_k – Hyperexponential distribution (k-th order)

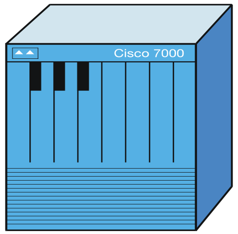
Modeling Example: Router

Router Model

router

packet arrivals

arrival times



packet departures

departure times

(waiting time + processing time)

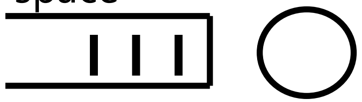
very simplified model

arrival process

arrival times

waiting space

service unit



waiting time

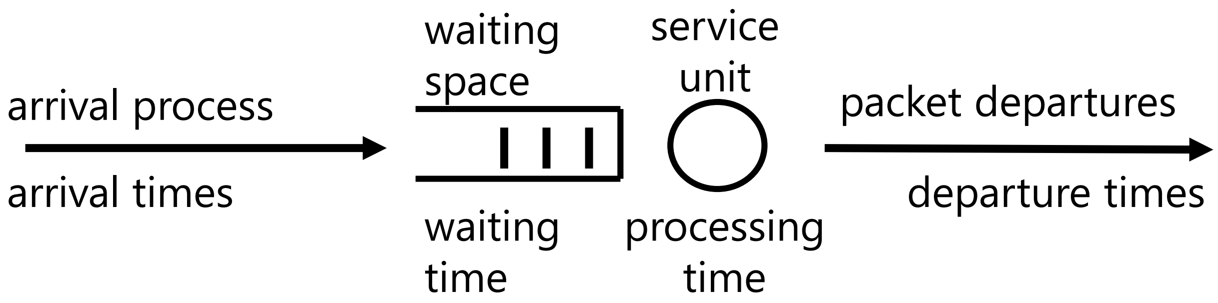
processing time

packet departures

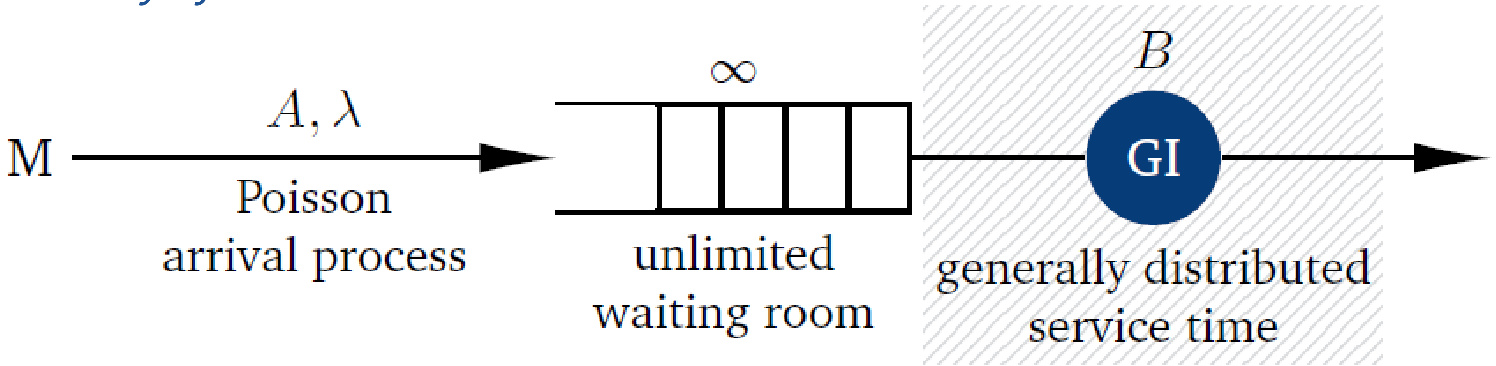
departure times

Router Model (f.)

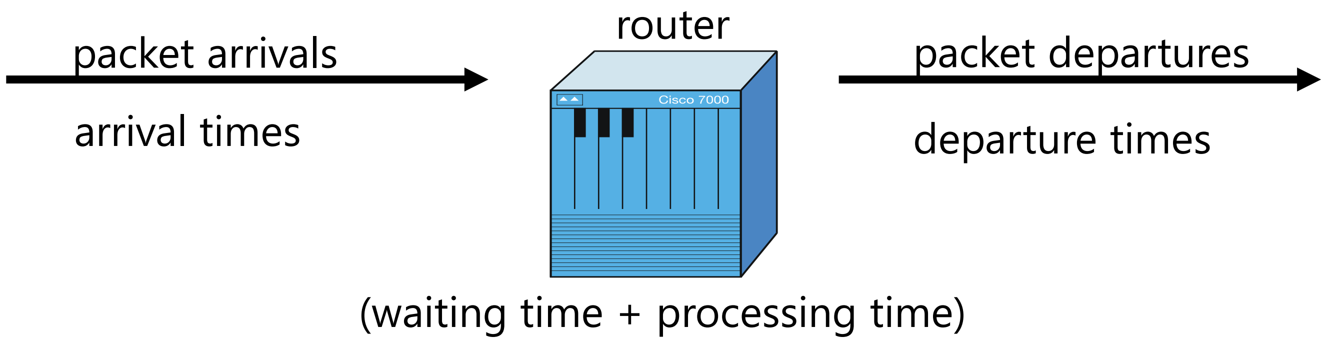
very simplified model



Example: M/GI/1 delay system



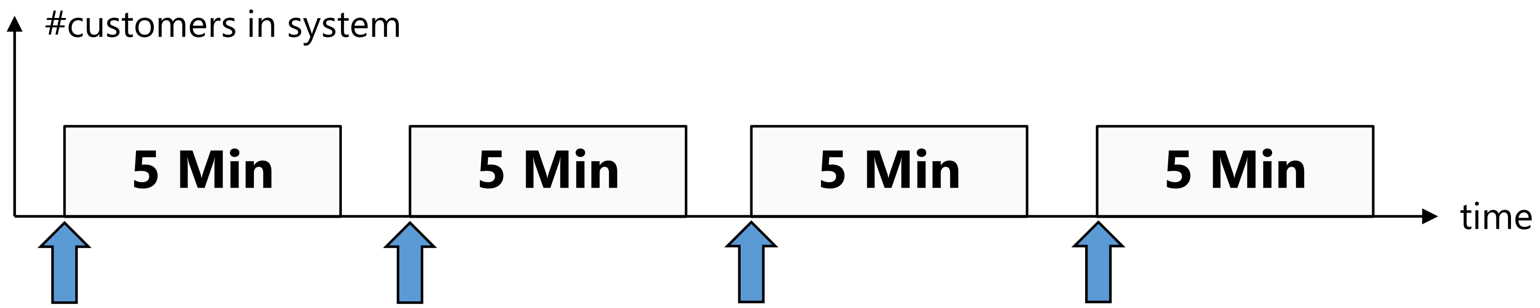
Performance Characteristics



- **Limiting resources**
  - processing speed of router
- **Performance measures**
  - queue length
  - throughput: #packets per time
  - sojourn / response time = waiting + processing time
  - utilization: fraction of time the router is busy (or idle)
- **Goal of model**
  - impact of router speed on buffer size (queue length), throughput and utilization
- **Queueing model**
  - M/GI/1-∞ delay system

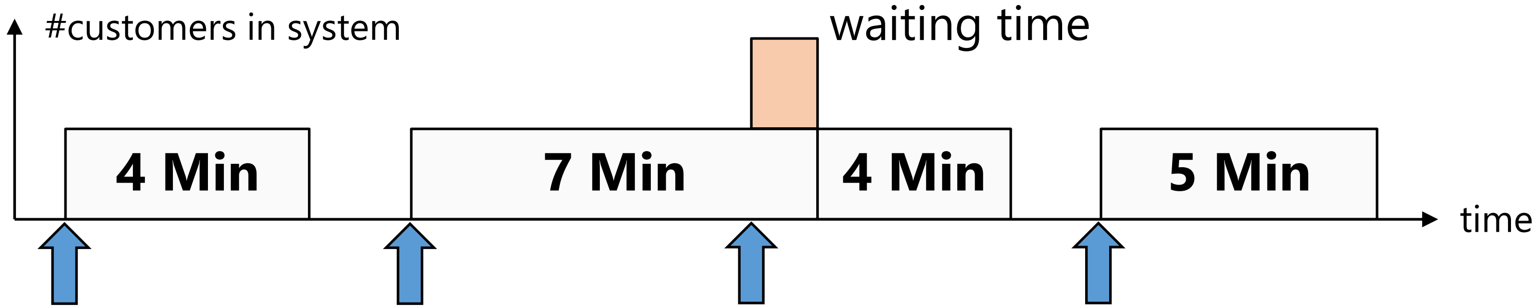
Different Delay Systems

- Periodic arrivals, identical service times, one server



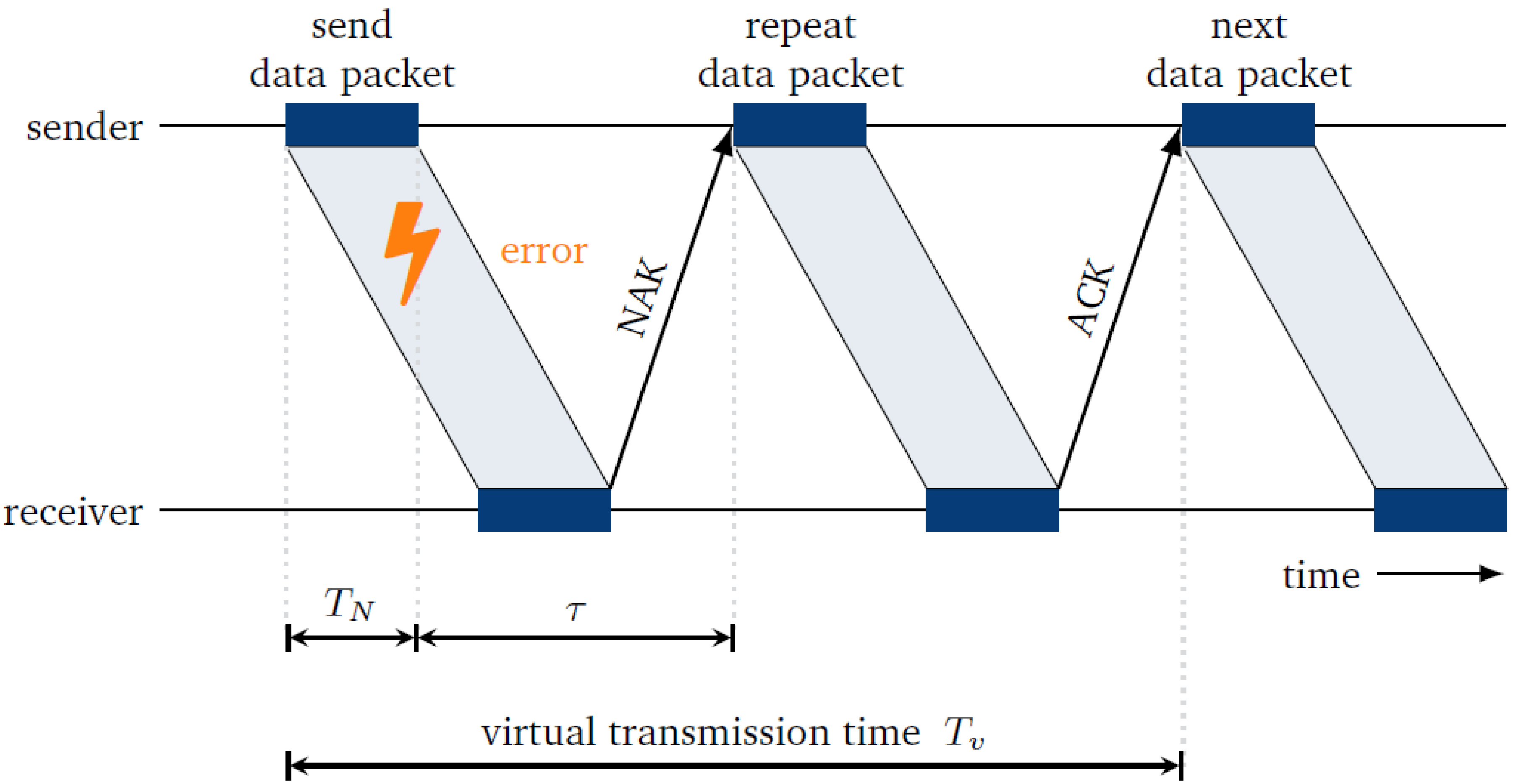
Kendall's notation?

- Periodic arrivals, random service times, one server

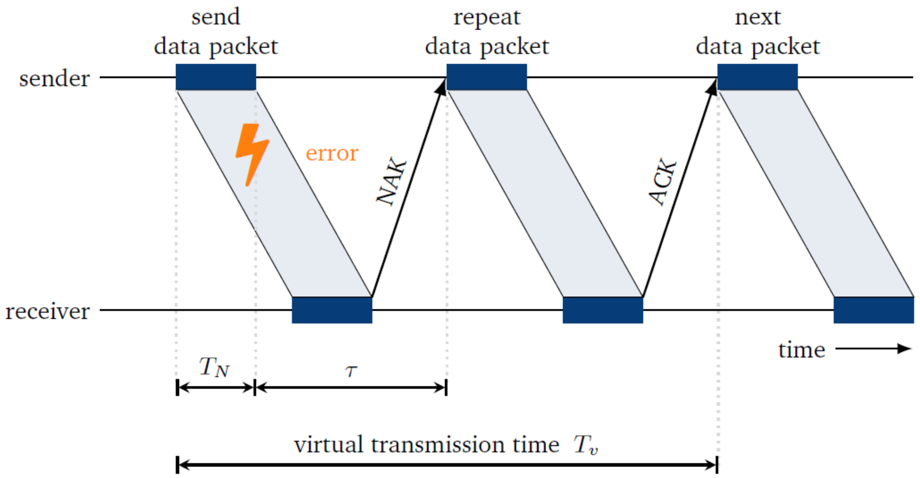


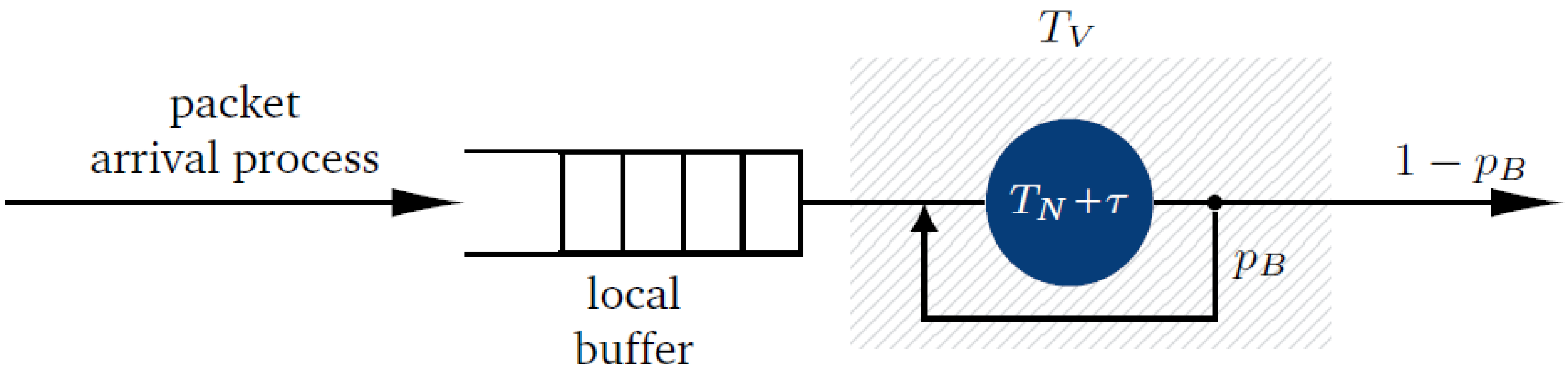
Modeling Example: Handshaking Protocol

Handshaking Protocol Illustration



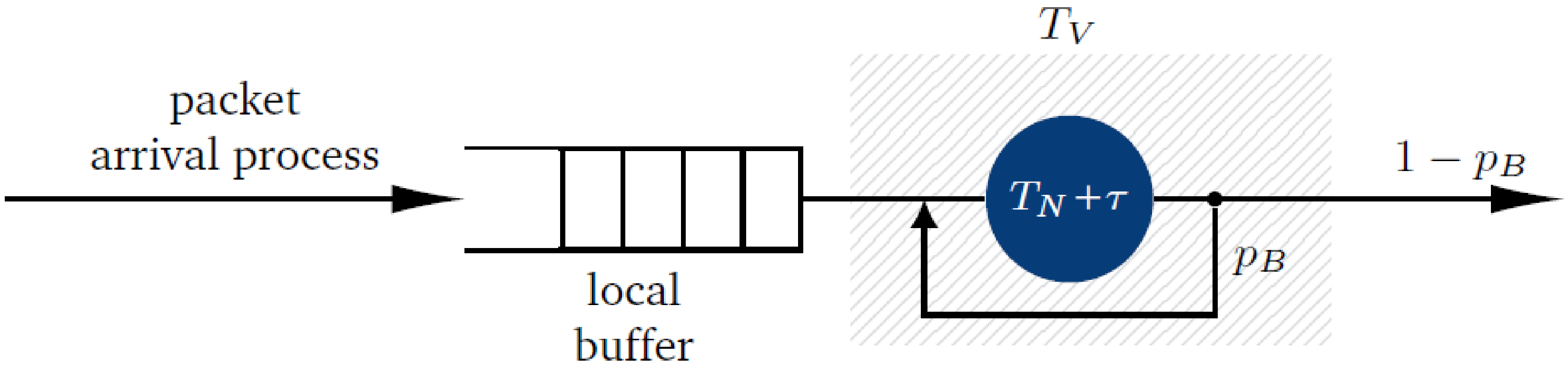
Traffic Model of the Handshaking Protocol



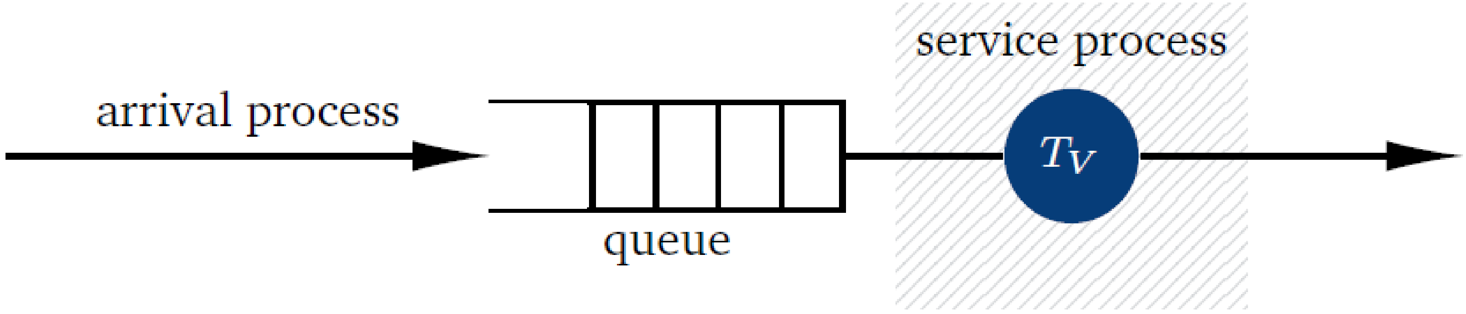


(a) Model of the handshaking protocol.

Traffic Model of the Handshaking Protocol (f.)



(a) Model of the handshaking protocol.



(b) Queueing model or traffic model.

Handshaking Protocol: Virtual Transmission Time

- Assume the following parameters
  - constant packet transmission time T_N
  - constant singal propagation delay τ from sender to receiver and back
  - packet error probability p_B for any packet

- What is the virtual transmission time?

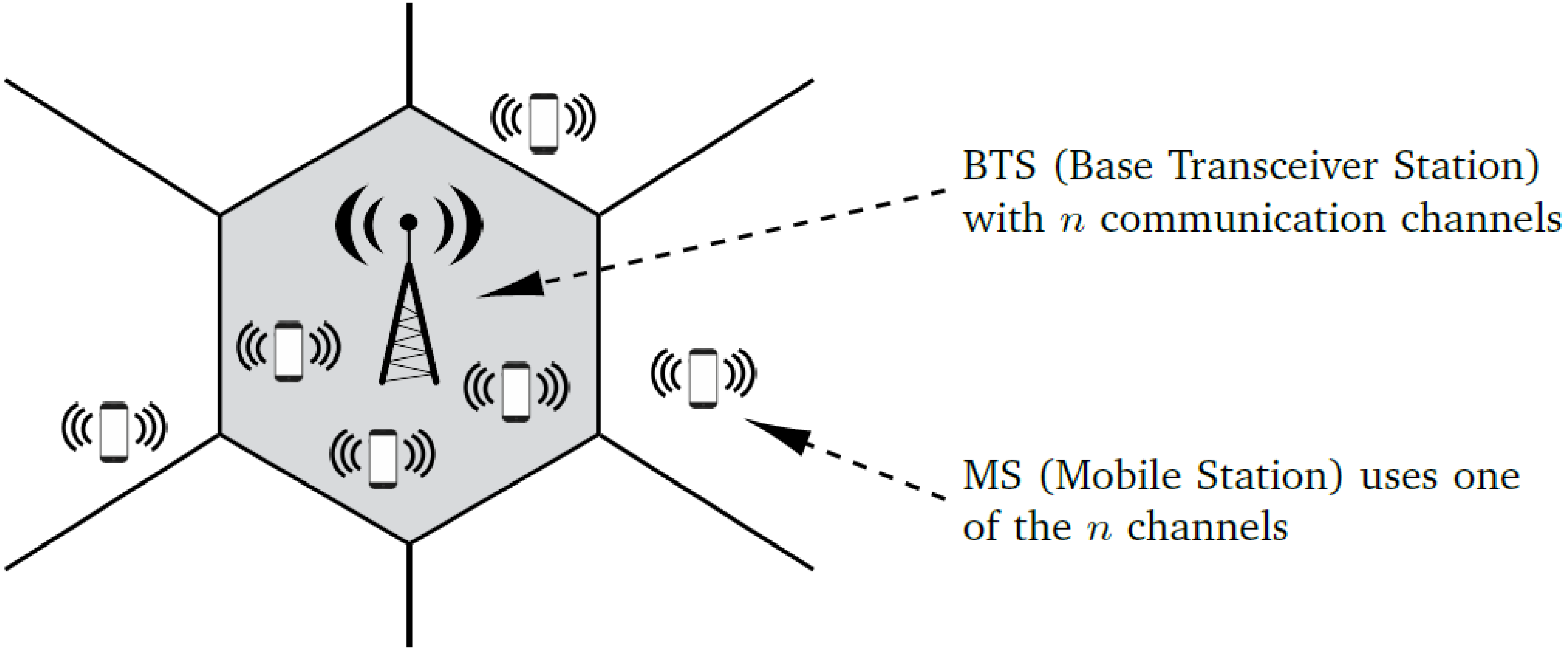
Handshaking Protocol Summary

- **Limiting resources**
  - processing of packets in the system, i.e., sending of packets over the error prone communication channel
- **Performance measures**
  - response time of the system (waiting time and virtual transmission time)
  - throughput per packet
- **Goal of model**
  - impact of packet error probability on performance measures
- **Queueing model**
  - GI/GI/1-∞ delay system

Modeling Example: Network Dimensioning in Mobile Networks

Network Dimensioning in Mobile Networks

- What is the arrival process? What is the service process?



Summary: Network Dimensioning in Mobile Networks

- **Limiting resources**
  - number of communication channels in the mobile cell
- **Performance measures**
  - blocking probability of customers' calls
- **Goal of model**
  - dimensioning of the number of required communication channels such that the blocking probability is below a QoS threshold for given parameters
- **Queueing model**
  - M/GI/n-0 loss system

Modeling Example: IoT Load Balancer

Superposition of Arrival Processes

- Kendall's notation is extended for the superposition of two or more input streams

- E.g. two arrival streams, e.g. from sensor 1 and sensor 2 with interarrival times A_1 and A_2

$$A_1+A_2/B/n\text{-}S$$

- For k independent arrival flows which are described by the same random variable A for the interarrival times

$$k\cdot A/B/n\text{-}S$$

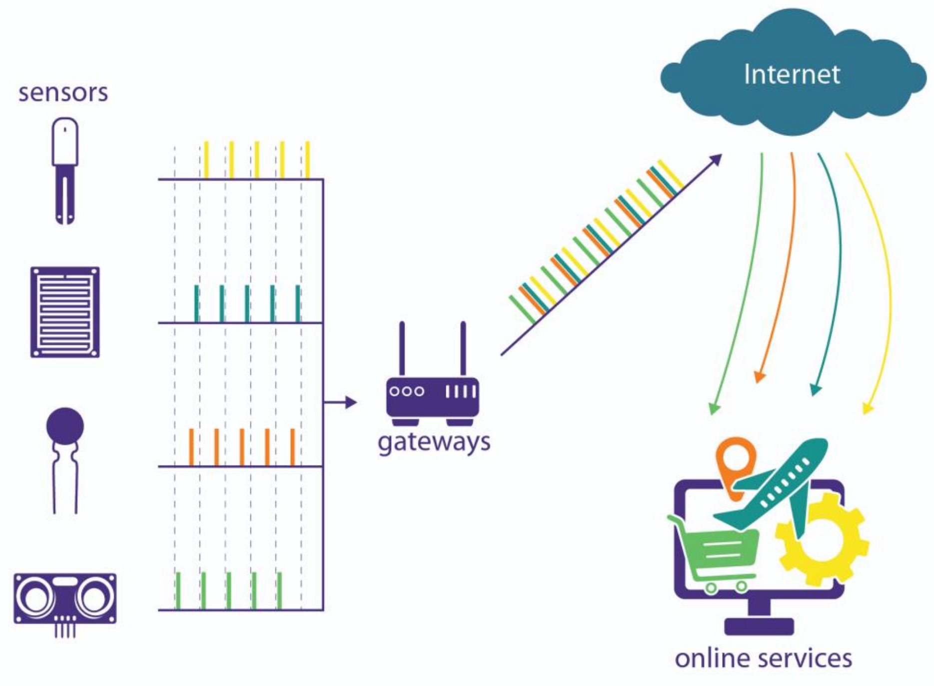
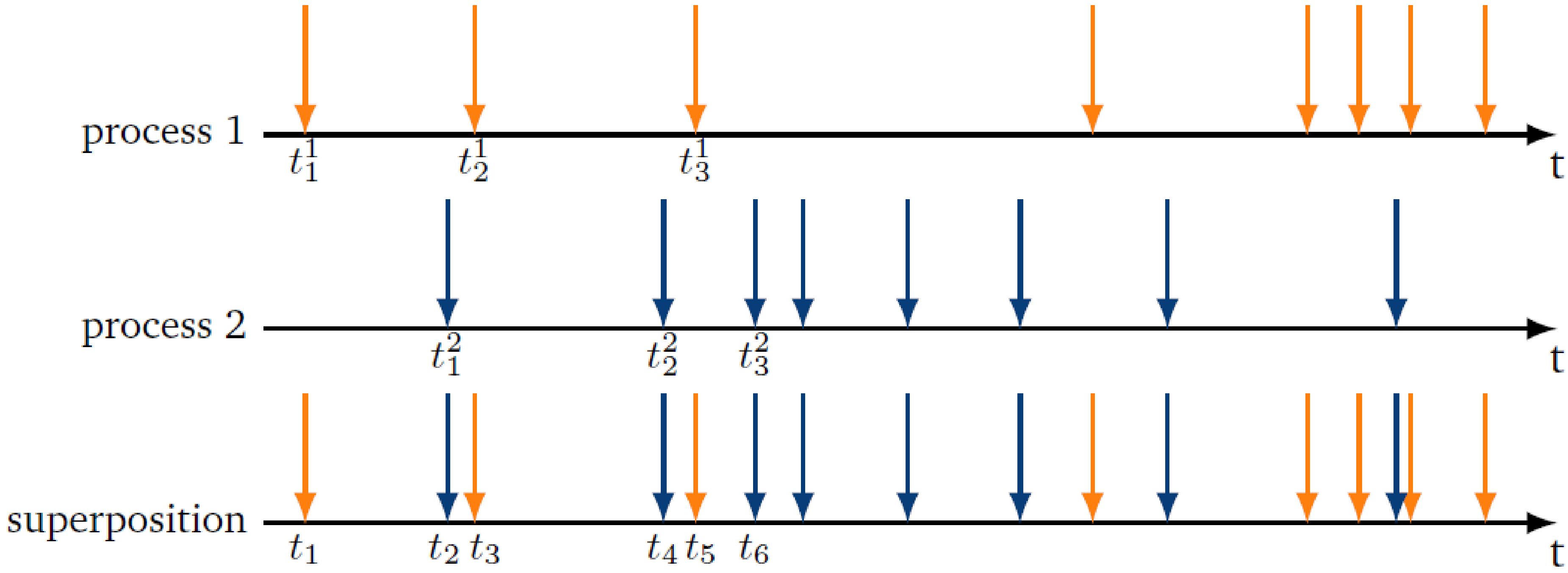
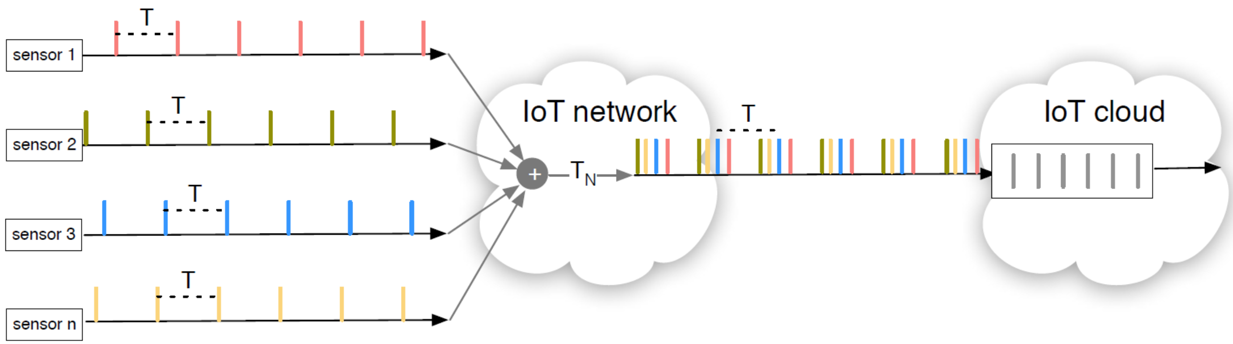


Illustration of Superposition of Arrival Processes



Example: IoT

- Sensors periodically send messages
- Deterministic processing time of (small) messages from the sensors
- Large number n of sensors



- How to model the IoT Load Balancer? Kendall's notation?
  - For single sensor $n = 1$?
  - For $n = 2$?
  - For large n ?

Summary: Network Dimensioning in Mobile Networks

- **Limiting resources**
  - IoT cloud load balancer
- **Performance measures**
  - waiting time or response time of IoT data packets at the load balancer
- **Goal of model**
  - dimensioning of the required processing power of the load balancer to keep the
  - mean waiting time below a predefined QoS threshold;
  - comparison of exact nD/D/1 model with the approximating M/D/1 queue
- **Queueing model**
  - nD/1/1 and M/D/1 delay system