



Module-3

IoT Processing Topologies and Types



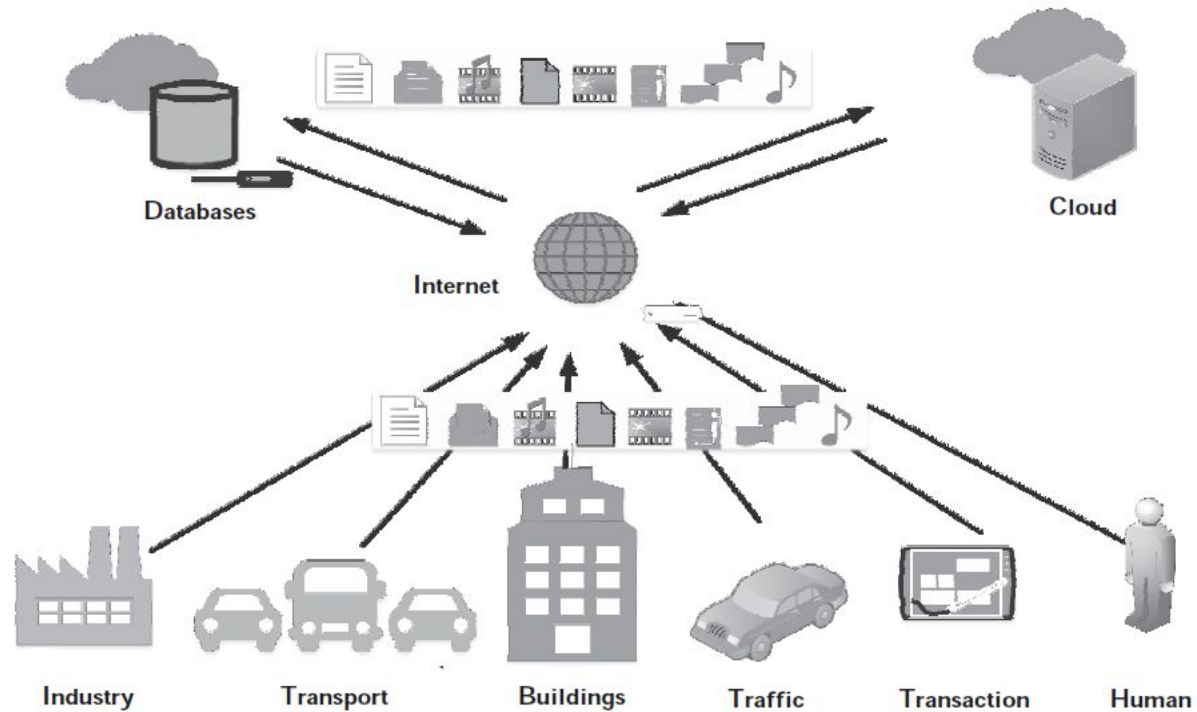
Contents

- Data Format
- Importance of Processing in IoT
- Processing Topologies
- IoT Device Design and Selection Considerations
- Processing Offloading



Data Format

- The Internet is a vast space where huge quantities and varieties of data are generated regularly and flow freely.
- As of January 2018, there are a reported 4.021 billion Internet users worldwide.
- The massive volume of data generated by this huge number of users
- Data-generating sources, non-human data generation sources such as sensor nodes and automated monitoring systems further add to the data load on the Internet
- Huge data volume is composed of a variety of data such as e-mails, text documents (Word docs, PDFs, and others), social media posts, videos, audio files, and images



The various data generating and storage sources connected to the Internet and the plethora of data types contained within it



Data can be broadly grouped into two types based on how they can be accessed and stored:

- 1) Structured data
- 2) unstructured data.



Structured data

- Text data that have a pre-defined structure
- Structured data are associated with relational database management systems (RDBMS)
- Primarily created by using length-limited data fields such as phone numbers, social security numbers, and other such information
- Even if the data is human or machine generated, these data are easily searchable by querying algorithms as well as human generated queries.
- This type of data is associated with flight or train reservation systems, banking systems, inventory controls, and other similar systems
- Structured Query Language (SQL) are used for accessing these data in RDBMS
- Structured data holds a minor share of the total generated data over the Internet.



Unstructured data

- All the data on the Internet, which is not structured, is categorized as unstructured.
- Data types have no pre-defined structure and can vary according to applications and data-generating sources.
- Human-generated unstructured data include text, e-mails, videos, images, phone recordings, chats, and others.
- Machine-generated unstructured data include sensor data from traffic, buildings, industries, satellite imagery, surveillance videos, and others.
- Data type does not have fixed formats associated with it, which makes it very difficult for querying algorithms to perform a look-up.
- Querying languages such as NoSQL are generally used for this data type.



Importance of Processing in IoT

- The vast amount and types of data flowing through the Internet necessitate the need for intelligent and resourceful processing techniques.
- This necessity has become even more crucial with the rapid advancements in IoT, which is laying enormous pressure on the existing network infrastructure globally
- The data to be processed is divided into three types based on the urgency of processing:
 - 1) Very time critical
 - 2) Time critical
 - 3) Normal



Very time critical

- Data from sources, which need immediate decision support, are deemed as very critical.
- Data have a very low threshold of processing latency, typically in the range of a few milliseconds.
- **Example:** flight control systems, healthcare, and other such sources

Data Processing:

- Exceptionally high
- The need for processing the data in place or almost nearer to the source is crucial in achieving the deployment success of such domains



Time critical

- Data from sources that can tolerate normal processing latency.
- Tolerate a latency of a few seconds fall in this category
- **Examples:** vehicles, traffic, machine systems, smart home systems, surveillance systems, and others

Data Processing:

- The transmission of data to be processed to remote locations/processors such as clouds or through collaborative processing



Normal

- Can tolerate a processing latency of a few minutes to a few hours
- **Example:** less data-sensitive domains such as agriculture, environmental monitoring, and others.

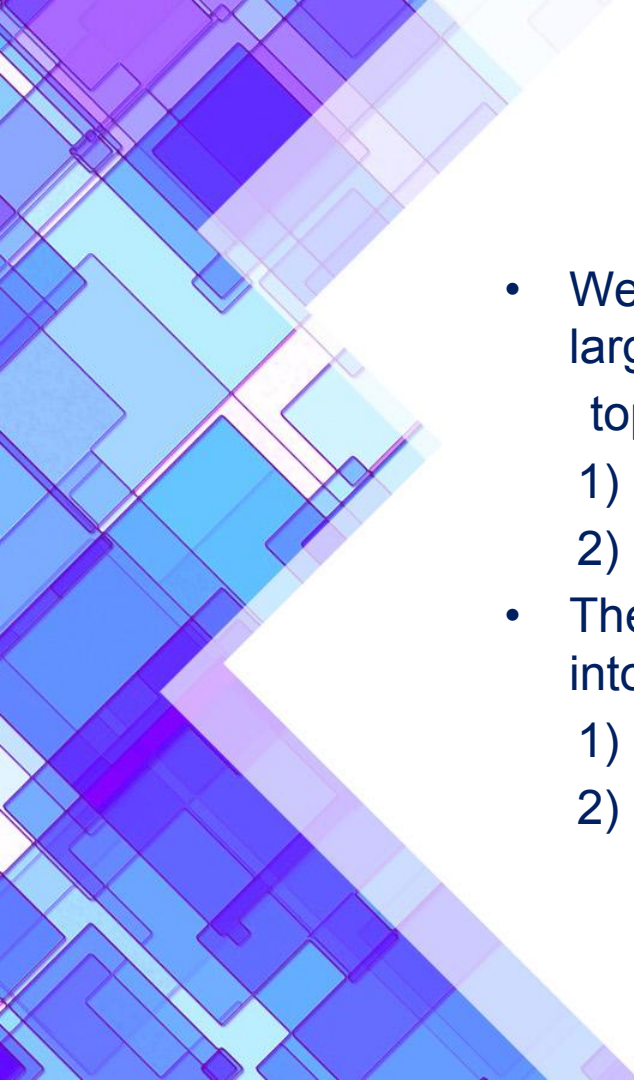
Data Processing

- Typically have no particular time requirements for processing urgently and are pursued leisurely as such.



Processing Topologies

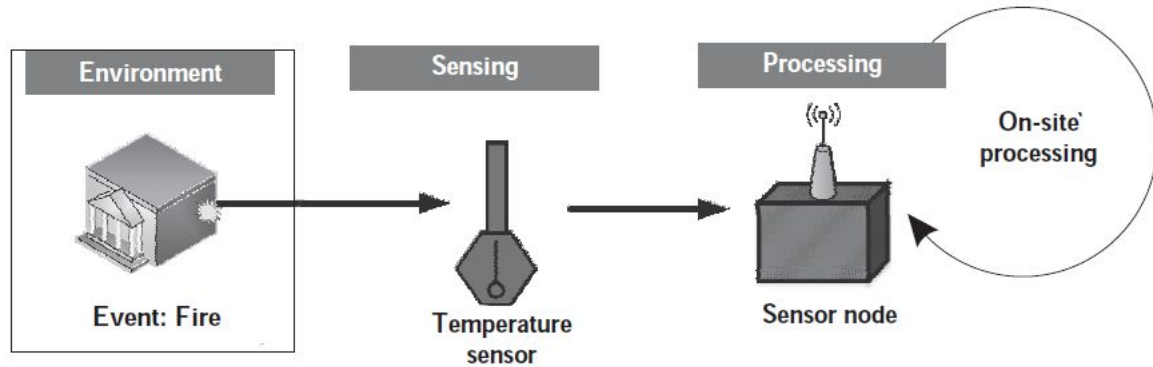
- The identification and intelligent selection of processing requirement of an IoT application are one of the crucial steps in deciding the architecture of the deployment.
- A properly designed IoT architecture would result in massive savings in
 - Network bandwidth
 - conserve significant amounts of overall energy in the architecture
 - while providing the proper and allowable processing latencies for the solutions associated with the architecture

- 
- We can divide the various processing solutions into two large topologies:
 - 1) On-site
 - 2) Off-site.
 - The off-site processing topology can be further divided into the following:
 - 1) Remote processing
 - 2) Collaborative processing.



On-site processing

- The data is processed at the source itself.
- This is crucial in applications that have a very low tolerance for latencies.
- These latencies may result from the processing hardware or the network
- Applications such as those associated with healthcare and flight control systems have a breakneck data generation rate.
- These additionally show rapid temporal changes that can be missed unless the processing infrastructure is fast and robust enough to handle such data

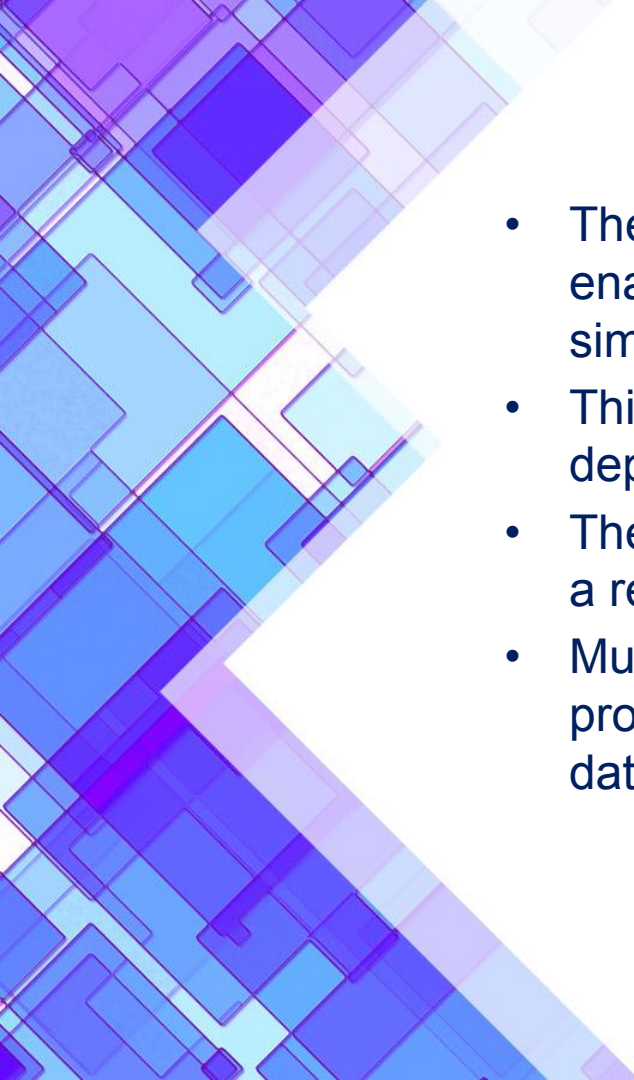


Event detection using an on-site processing topology



Off-site processing

- As opposed to the on-site processing paradigms, allows for latencies.
- It is significantly cheaper than on-site processing topologies.
- Difference in cost is mainly due to the low demands and requirements of processing at the source itself
- The sensor nodes are not required to process data on an urgent basis, so having a dedicated and expensive on-site processing infrastructure is not sustainable for large-scale deployments typical of IoT deployments.
- The sensor node is responsible for the collection and framing of data that is eventually to be transmitted to another location for processing.

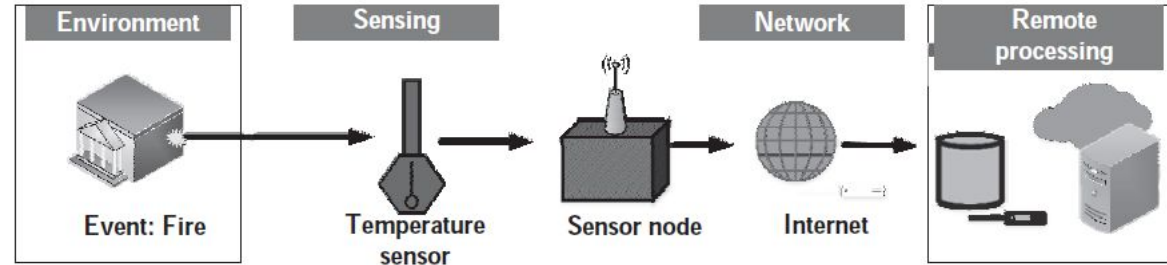
- 
- The off-site topology has a few dedicated high-processing enabled devices, which can be borrowed by multiple simpler sensor nodes to accomplish their tasks.
 - This arrangement keeps the costs of large-scale deployments extremely manageable
 - The data from these sensor nodes is transmitted either to a remote location or to multiple processing nodes.
 - Multiple nodes can come together to share their processing power in order to collaboratively process the data.



Remote processing

- The most common processing topologies prevalent in present-day IoT solutions.
- It encompasses sensing of data by various sensor nodes; the data is then forwarded to a remote server or a cloud-based infrastructure for further processing and analytics.
- The processing of data from hundreds and thousands of sensor nodes can be simultaneously offloaded to a single, powerful computing platform;
- This results in massive cost and energy savings by enabling the reuse and reallocation of the same processing resource while also enabling the deployment of smaller and simpler processing nodes at the site of deployment.

- Setup also ensures massive scalability of solutions, without significantly affecting the cost of the deployment.

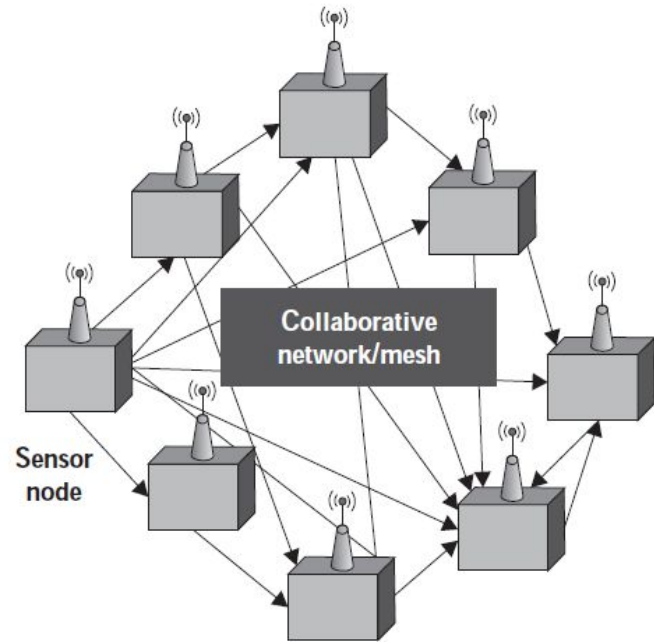
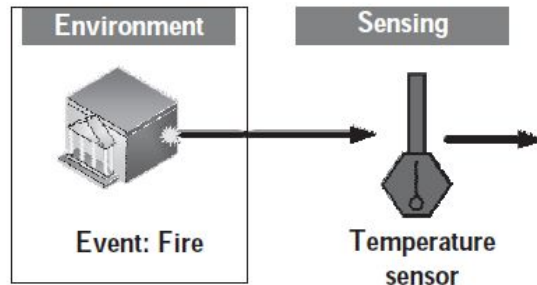


Event detection using an o-site remote processing topology



Collaborative processing

- Processing topology typically finds use in scenarios with limited or no network connectivity, especially systems lacking a backbone network.
- Quite economical for large-scale deployments spread over vast areas, where providing networked access to a remote infrastructure is not viable.
- The simplest solution is to club together the processing power of nearby processing nodes and collaboratively process the data in the vicinity of the data source itself.
- Reduces latencies due to the transfer of data over the network.
- It conserves bandwidth of the network, especially ones connecting to the Internet.



Event detection using a collaborative processing topology



IoT Device Design and Selection Considerations

- The main consideration of minutely defining an IoT solution is the selection of the processor for developing the sensing solution.
- This selection is governed by many parameters that affect the usability, design, and affordability of the designed IoT sensing and processing solution.
- The main factor governing the IoT device design and selection for various applications is the Processor.
- Other considerations are
 - Size
 - Energy
 - Cost
 - Memory
 - Processing power
 - I/O rating
 - Add-ons



Size

- The crucial factors for deciding the form factor and the energy consumption of a sensor node.
- Larger the form factor, larger is the energy consumption of the hardware
- Large form factors are not suitable for a significant bulk of IoT applications, which rely on minimal form factor solutions (e.g., wearables).



Energy

- The energy requirements of a processor is the most important deciding factor in designing IoT-based sensing solutions.
- Higher the energy requirements, higher is the energy source (battery) replacement frequency.
- This principle automatically lowers the long-term sustainability of sensing hardware, especially for IoT-based applications.

Cost

- The cost of a processor, besides the cost of sensors, is the driving force in deciding the density of deployment of sensor nodes for IoT-based solutions.
- Cheaper cost of the hardware enables a much higher density of hardware deployment by users of an IoT solution.
- For example, cheaper gas and fire detection solutions would enable users to include much more sensing hardware for a lesser cost.



Memory

- The memory requirements (both volatile and non-volatile memory) of IoT devices determines the capabilities the device can be armed with.
- Features such as local data processing, data storage, data filtering, data formatting, and a host of other features rely heavily on the memory capabilities of devices.
- Devices with higher memory tend to be costlier for obvious reasons.



Processing power

- Processing power is vital in deciding what type of sensors can be accommodated with the IoT device/node, and what processing features can integrate on-site with the IoT device.
- The processing power also decides the type of applications the device can be associated with.
- Applications that handle video and image data require IoT devices with higher processing power as compared to applications requiring simple sensing of the environment.



I/O rating

- The input–output (I/O) rating of IoT device, primarily the processor, is the deciding factor in determining the circuit complexity, energy usage, and requirements for support of various sensing solutions and sensor types.
- Newer processors have a lesser I/O voltage rating of 3.3 V, as compared to 5 V for the somewhat older processors.
- This translates to requiring additional voltage and logic conversion circuitry to interface legacy technologies and sensors with the newer processors.
- Despite low power consumption due to reduced I/O voltage levels, this additional voltage and circuitry not only affects the complexity of the circuits but also affects the costs.




Add-ons

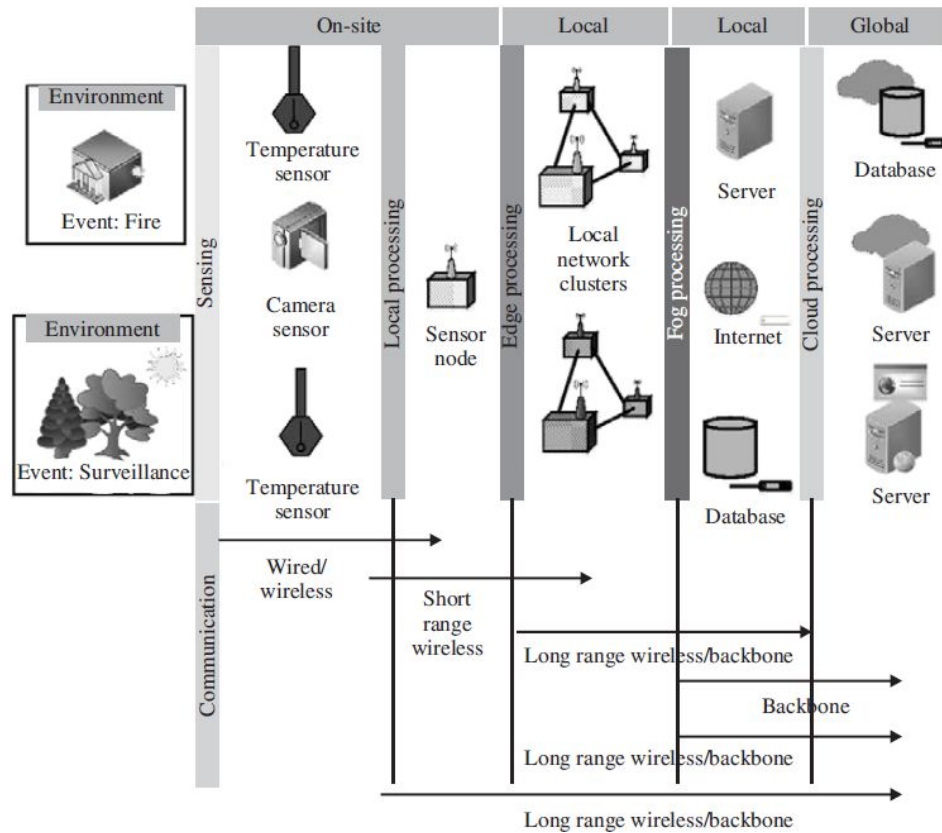
- The support of various add-ons a processor or for that matter, an IoT device provides, such as
 - analog to digital conversion (ADC) units,
 - in-built clock circuits,
 - connections to USB and ethernet,
 - inbuilt wireless access capabilities, and
 - others
- Helps in defining the robustness and usability of a processor or IoT device in various application scenarios.
- These add-ons also decides how fast a solution can be developed, especially the hardware part of the whole IoT application.



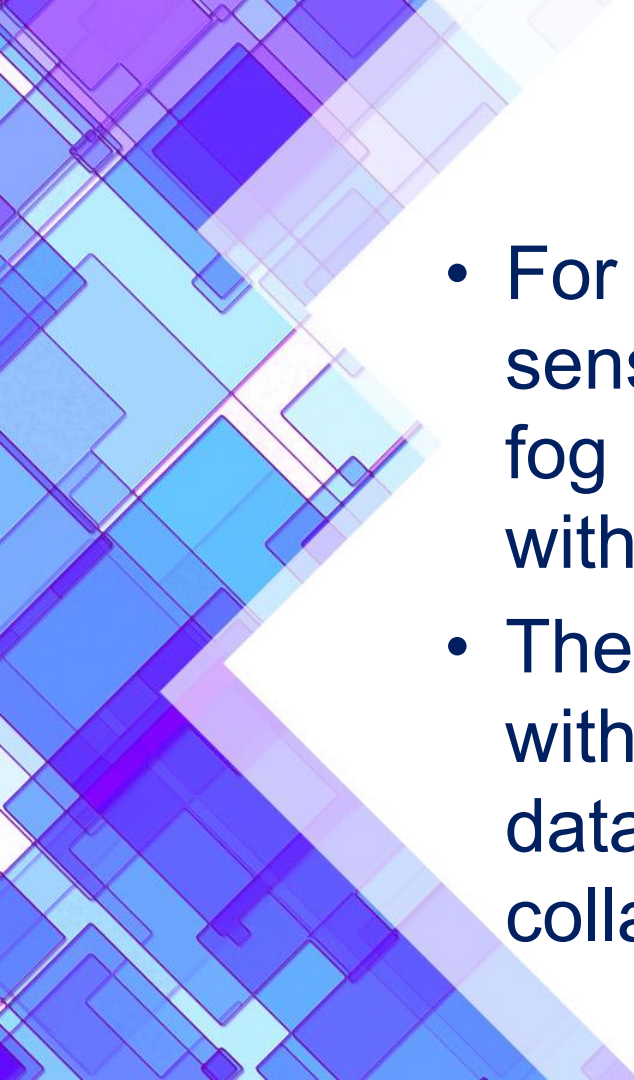
Processing Offloading

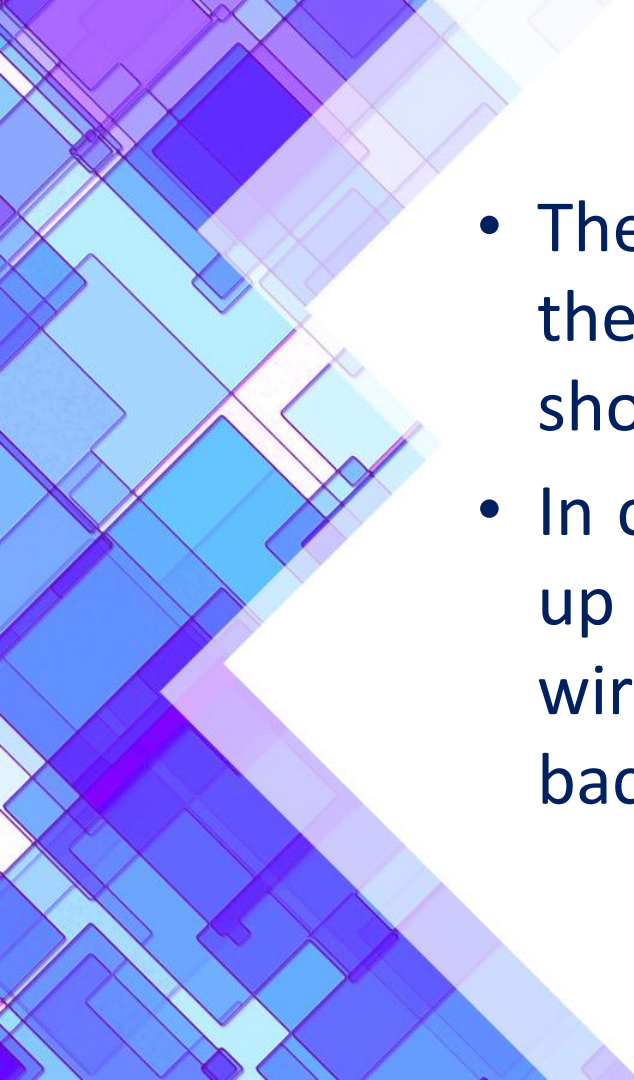
- Processing offloading is important for the development of densely deployable, energy-conserving, miniaturized, and cheap IoT-based solutions for sensing tasks
- Figure shows the typical outline of an IoT deployment with the various layers of
- processing that are encountered spanning vastly different application domains—from as near as sensing the environment to as far as cloud-based infrastructure.


- 
- Starting from the primary layer of sensing, multiple sensing types tasked with detecting an environment are considered (fire, surveillance, and others).
 - The sensors enabling these sensing types are integrated with a processor using wired or wireless connections (mostly, wired).
 - In the event that certain applications require immediate processing of the sensed data, an on-site processing topology is followed



The various data generating and storage sources connected to the Internet and the large amount of data types contained within it

- 
- For off-site processing, data from the sensing layer can be forwarded to the fog or cloud or can be contained within the edge layer
 - The edge layer makes use of devices within the local network to process data that which is similar to the collaborative processing topology


- 
- The devices within the local network, till the fog, generally communicate using short-range wireless connections.
 - In case the data needs to be sent further up the chain to the cloud, long-range wireless connection enabling access to a backbone network is essential.


- 
- Fog-based processing is still considered local because the fog nodes are typically localized within a geographic area and serve the IoT nodes within a much smaller coverage area as compared to the cloud.
 - Fog nodes, which are at the level of gateways, may or may not be accessed by the IoT devices through the Internet.




Offload location

- The choice of offload location decides the applicability, cost, and sustainability of the IoT application and deployment.
- We distinguish the offload location into four types:
 - Edge
 - Fog
 - Remote Server
 - Cloud


- 
- **Edge:** Offloading processing to the edge implies that the data processing is facilitated to a location at or near the source of data generation itself.
 - Offloading to the edge is done to achieve aggregation, manipulation, bandwidth reduction, and other data operations directly on an IoT device .

- 
- **Fog:** Fog computing is a decentralized computing infrastructure that is utilized to conserve network bandwidth, reduce latencies, restrict the amount of data unnecessarily flowing through the Internet, and enable rapid mobility support for IoT devices.



- **Remote Server:** A simple remote server with good processing power may be used with IoT-based applications to offload the processing from **resource constrained IoT devices**.

Rapid scalability may be an **issue** with remote servers, and they may be costlier and hard to maintain in comparison to solutions such as the cloud .

- 
- **Cloud:** Cloud computing is a configurable computer system, which can get access to configurable resources, platforms, and high-level services through a shared pool hosted remotely.
 - A cloud is provisioned for processing offloading so that processing resources can be rapidly provisioned with minimal effort over the Internet, which can be accessed globally.



Offload Decision Making

- The choice of where to offload and how much to offload is one of the major deciding factors in the deployment of an offsite-processing topology-based IoT deployment architecture.
- The decision making is generally addressed considering data generation rate, network bandwidth, the criticality of applications, processing resource available at the offload site, and other factors
- Some Approaches are
 - Naive Approach
 - Bargaining based Approach
 - Learning based Approach



Naïve Approach

- Typically a hard approach, without too much decision making.
- It can be considered as a rule-based approach in which the data from IoT devices are offloaded to the nearest location based on achievement of certain offload criteria.
- Although easy to implement, this approach is never recommended, especially for dense deployments, or deployments where the data generation rate is high or the data being offloaded is complex to handle (multimedia or hybrid data types).
- Generally, statistical measures are consulted for generating the rules for offload decision making.



Bargaining based Approach

- Although a bit processing-intensive during the decision making stages, enables the lessening of network traffic congestion, enhances service QoS (quality of service) parameters such as bandwidth, latencies, and others.
- At times, while trying to maximize multiple parameters for the whole IoT implementation, in order to provide the most optimal solution or QoS, not all parameters can be treated with equal importance.
- Bargaining based solutions try to maximize the QoS by trying to reach a point where the qualities of certain parameters are reduced, while the others are enhanced.
- This measure is undertaken so that the achieved QoS is collaboratively better for the full implementation rather than a select few devices enjoying very high QoS.
- Game theory is a common example of the bargaining based approach.
- This approach does not need to depend on historical data for decision making purposes.



Learning based Approach

- The learning based approaches generally rely on past behavior and trends of data flow through the IoT architecture.
- The optimization of QoS parameters is pursued by learning from historical trends and trying to optimize previous solutions further and enhance the collective behavior of the IoT implementation.
- The memory requirements and processing requirements are high during the decision making stages.
- The most common example of a learning based approach is machine learning.



Offloading considerations

- Few offloading parameters which need to be considered while deciding upon the offloading type to choose.
- These considerations typically arise from the nature of the IoT application and the hardware being used to interact with the application.
- Some of these parameters are as follows.
 - Bandwidth
 - Latency
 - Criticality
 - Resources
 - Data Volume



Bandwidth

- The maximum amount of data that can be simultaneously transmitted over the network between two points is the bandwidth of that network.
- The bandwidth of a wired or wireless network is also considered to be its data-carrying capacity and often used to describe the data rate of that network.

Latency

- It is the time delay incurred between the start and completion of an operation.
- In the present context, latency can be due to the network (network latency) or the processor (processing latency).
- In either case, latency arises due to the physical limitations of the infrastructure, which is associated with an operation.
- The operation can be data transfer over a network or processing of a data at a processor.



Criticality

- It defines the importance of a task being pursued by an IoT application.
- The more critical a task is, the lesser latency is expected from the IoT solution.
- For example, detection of fires using an IoT solution has higher criticality than detection of agricultural field parameters.
- The detection of fire requires a response time in the tune of milliseconds, whereas the agricultural field can be addressed within hours or even days.



Resources

- It signifies the actual capabilities of an offload location.
- These capabilities may be the processing power, the suite of analytical algorithms, and others.



Data volume

- The amount of data generated by a source or sources that can be simultaneously handled by the offload location is referred to as its data volume handling capacity.
- Typically, for large and dense IoT deployments, the offload location should be robust enough to address the processing issues related to massive data volumes.



Thank you