Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 11, November-2020

IoT Applications on 5G Edge

Shelley Sam Varughese
Student: Department of Computer Science
Musaliar College of Engineering and Technology,
Affiliated to APJ Abdul Kalam Technological University,
Pathanamthitta, Kerala, India.

Prof. Deepa Thomas
Assistant Professor: Department of Computer Science
Musaliar College of Engineering and Technology,
Affiliated to APJ Abdul Kalam Technological University,
Pathanamthitta, Kerala, India.

Dr. Manikandan L C
Head of the Department: Department of Computer Science
Musaliar College of Engineering and Technology,
Affiliated to APJ Abdul Kalam Technological University,
Pathanamthitta, Kerala, India.

***Abstract*—: Internet of things (IOT) primarily consists of a system that connects to the internet. The IoT related concepts like self-driving cars, smart cities, e-health care, etc. have a ubiquitous presence now. These applications require higher data-rates, larger bandwidth, increased capacity, low latency and high throughput. The key for the future IoT system is the shift from a static architecture to a dynamically self-organizing and evolving one. Devices of extremely varying capabilities needs to collaborate and access all necessary information to ensure their optimal work, while keeping the flexibility in the network configuration. The supporting infrastructure should allow the connected devices to interact with the most convenient node in the network and should be able to optimize the resource consumption, without compromising QoS. The main idea is to summarize the process of building an offloading framework for the arbitrary task of IoT devices. 5G cellular networks provide the enabling technologies for deployment of the IoT technology everywhere and anywhere. Another contribution is the collection of requirements like need for Edge Computing, Offloading paradigms and an underlying infrastructure for the IoT networks and computing systems, which will be supported by means of Clustering, 5G mobile networking standards, and Artificial Intelligence.**

Keywords— IoT networks, Edge Computing,5G, Artificial Intelligence.

I. INTRODUCTION

IoT applications handles informations from a number of heterogeneous devices. 5g is the foundation for realizing full potential of IoT. In the past generation IoT devices which were considered as small, external hardwares with limited resources, like sensors which resides at the edge of the involved network infrastructure. This assumption made the main role of the IoT devices to blindly transmit sensed data or to react to environmental changes to some extent.

Fig 1: Block diagram showing IoT based smart applications

The limitation on the computing resources on the IoT devices, made practice is of offloading tasks of various applications to the computing systems with resources like data centres in the cloud. However, the main drawbacks of the offloading methods were high latency and network congestion in the infrastructure. This issue gave rise to the paradigm of Edge Computing, the idea was to support the devices with a cloud closer to the edge of the network. This appeared as a solution. However, adding Edge resources complicated the management of the network because multiple devices will be contending them.

Furthermore, the recent evolution of IoT brought more and more devices which were not simple sensors or transmitters. It provided a limited execution environment. This opened up a huge opportunity to utilize this previously unused processing power in order to offload custom application logic directly to these edge devices. In this very complex scenario, it is an essential question of how to balance the tasks and the resources available in such a way that would profit from the added capabilities of the IoT devices without compromising the final performance.

The key question in future of IoT networks then was how to enable devices of extremely varying capabilities to collaborate and access all information necessary for ensuring the optimal work while keeping the flexibility in the network configuration. An important issue in relation with collaboration of IoT devices is connectivity, since the continuously growing number of devices could generate congestions in the communication channels.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 11, November-2020

The last change in the IoT is mobility. We would have to consider many kinds of objects. The definition of" Things" is very broad, consisting from smart phones to even smart cars. Things are actually any physical objects which have a real-life presence. Such objects could be installed on moving vehicles will be mobile themselves.

For all presented scenarios, there will be dynamic changes in configuration at the edge of the network. This happens very frequently. For example, the connected devices may be physically moving and the network might need to balance the resources or will have to reallocate them to achieve system faults tolerance. These continuous changes will require the IoT networks to be able to reorganize itself in such a way that optimizes the resource consumption like bandwidth, storage and power. And allows the connected devices to interact with the most convenient node in the network, without compromising the Quality of the Service (QoS). Simply copying the whole applications in every node which requires them cannot be scalable or a maintainable solution and offloading framework is still needed.

The answer to the connectivity question of the ever-increasing number of devices is a solution which is already applied for Wireless Sensor Networks (WSNs) and it is to form groups of devices and to manage its connections in a collaborative way.

A network of such devices could profit from the 5th Generation (5G) mobile network standards and its infrastructures in order to achieve the goals which is otherwise unattainable with only Edge computing and offloading.

The network should be able to adapt to fluctuations of resource load. As the 5G networks are more complex architectures than their predecessors, the number of configuration variable makes it very difficult to apply the deterministic adjustment approaches. Due to this, we could believe that the IoT framework will also benefit from the Artificial Intelligence (AI) based technologies which are designed precisely to cope with these challenges.

The main effort is to summarize and redefine the requirements for this new IoT networks and computing systems, the need for a platform on top of 5G edge computing and computational offloading paradigms. Compared to previous works we have go into more details regarding the possible technical solution that can be used to satisfy the requirements of this IoT platform, in particular we could give the overview of the different choices for modelling and partitioning the application and we could individuate some progress in clustering and AI that can support the development of such complex solutions.

In the effort to summarize the processes needed for building an offloading framework for arbitrary task of the IoT devices we specified five main parts:

- Discovery and modelling
- Planning and optimization
- Execution
- Monitoring and performance maintenance
- Learning and predicting

II. BACKGROUND

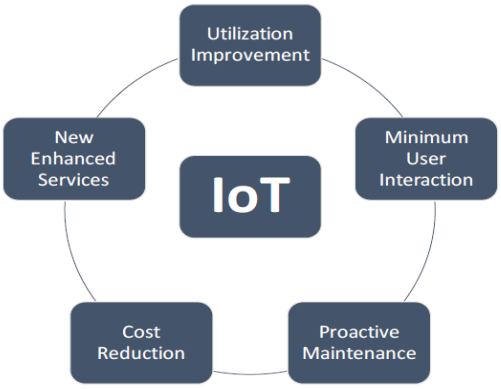

Fig 2: Block diagram showing IoT benefits as the cutting-edge technology

In this section the basic definitions of the involved technologies and paradigms are tried to summarize the mainstream directions of the literature.

A. EDGE COMPUTING

As the number of mobile devices and services that requires computational or storage capabilities that significantly exceeded their own capacities, the paradigm of Cloud Computing (CC) rose and gained a continuously increasing importance. The concept of cloud computing is based on Data Centres (DC), which are capable of managing the processing and storage requirements of the tasks involving very large data. Moreover, Data Centre Networks (DCNs) are built up by connecting the data centres using optical cables. Due to the extremely low internal communication costs it is appeared as a single entity in the outside world. When a problem occurs that will outrun the available resources one can offload the data and the code to the cloud and after the computations are done the results are given back.

The paradigm of Cloud computing has thus given a solution to the scalability related problems due to inefficient resources. There is a significant latency or congestion in the network due to the code and data migration. Location unawareness CC paradigm is the cause of the problem and it will get more severe time as more and more (semi-)intelligent devices attempts to connect to DCNs.

The concept of Edge Computing emerged and it leveraged the storage and computation capabilities of the edge devices which are connected to the Internet and was meant to be an intermediate layer between the devices. They are able to handle the subset of requests is usually sent to the cloud, but didn't need its real involvement due to the diminished resource requirement or because of the no need for the DC to be involved. Thus, the computation load of DCs is reduced to some extent. The latency of responses when an application needs to have real-time or almost real-time responses also have been reduced. Moreover, due to its high availability and geo-distributed nature, the Edge layer is appropriate to handle the challenges of mobility; for e.g., serve the requests of moving users like in the cases of autonomous cars or streaming and real-time gaming.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 11, November-2020

1) ***Comparison of Edge Computing Implementations: Fog Computing, Cloudlet and Mobile Edge Computing:*** In the following *the* distinction of Edge Computing into three categories.

a) Fog Computing: In this, Fog Computing Nodes can be placed at any point of the architecture. They are highly heterogeneous and can be built on various devices such as routers, switches, IoT gateways etc. Heterogeneity of devices leaded to the ability for working with different protocols and with non-IP-based technologies that is in communication between fog computing nodes and the end devices. Since heterogeneity of the edge naturally stays hidden from the user devices. Fog computing system exposes a uniform interface containing storage and computational services. Also, the monitoring security and device management facilities are exposed. On top of this abstraction layer, the orchestration layer organizes resource allocations according to the users' requests.

b) Cloudlet: Cloudlet can be defined as a trusted set of computers which have a good connection to the Internet and their resources are made available to nearby mobile devices. The Cloudlet runs a virtual machine which is capable of arranging resources to the connected users near real-time over the WLAN network within one-hop distance and high bandwidth. Above provisioning the infrastructure, Cloudlet architecture provides a middle framework support to component-based applications designed with a focus on applications with strict real-time requirements such as augmented reality.

c) Mobile Edge Computing: It is responsible for bringing computational resources and storage to the edge of Radio Access Network which enables us to reduce the latency and improve location awareness. A clear choice for deploying the mobile edge computing nodes is to collocate them with the Radio Network Controller or a macro base-station. Servers of those node runs multiple mobile edge computing host instances with their own storage and computational resources. An orchestrator will be in charge for monitoring the hosts and state of the network. It keeps track of which host offer what services, tracks the available resources as well as information on devices which are connected to the servers including location and routing information.

B. 5G

The standardization process of Fifth Generation of mobile network is still going on. Its main characteristics are more or less publicly available. The technology will be based on:

1) *Economic fibre-like radio access* reaching data rates beyond 10 Gb/s, by the usage of higher frequency bands above 6 GHz and related technologies.

2) *Network Function Virtualization (NFV)*, allows implementing specific network functions in software running on generic hardware without the need for costly hardware-specific machines. Reducing implementation, management and operational costs. It allows the reuse and sharing of the same functionality between customers.

3) *Software Defined Networking (SDN)*, will allow the control of network resources to be opened to third parties and flexibility to accommodate demanding professional grade applications.

Advantages of 5G over 4G are:

1) Very low latency of 1ms which was about 10 to 20ms for 4G.

2) Serving about 1 million of devices per square Km which was about 1000 device per square km for 4G.

3) Fast deployment of new services in 1-hour time which was done in days with current technology.

C. OFFLOADING AND PARTITIONING

Offloading is the act of transferring a certain computing tasks to an external platform. The main motive is to augment the mobile systems' capabilities through code migration which enables the systems to improve their performance and save energy. Offloading can be very useful in case of computation and data intensive applications such as AI and object tracking.

Offloading will occur when it is beneficial for the mobile application. In most cases offloading occurs when outsourcing the computation could improve the response time or save energy. Calculating this is very complicated and depends upon several parameters. Whether and what to offload are determined upon the most frequently based on parameters of the network, hosting device, and the cloud infrastructures. This will include bandwidths, server speeds, available memory etc. Computing benefits in time. For e.g. it should include time lost for the transmission and time gained on the faster edge computation. In the perspective of total energy consumption, the most important factor is that the edge/cloud is more energy efficient than the smaller device which have energy constraints (battery). All of these is taken into account with the priority specified. The priority is based on the" optimization" criteria. The heavy loaded network could inhibit offloading.

The technique of splitting up the application into separate components, while preserving the semantics of the original application is known as Application Partitioning (AP). The main objective of partitioning algorithms is to divide the code into different logical units or tasks (candidates for offloading) and to clearly specify the interfaces of interaction between them.

Partitioning algorithms can be classified according to different parameters like task granularity, application model, partitioning objective, language support, profiler used, analysis technique, allocation decision, and usage of annotation (automatic or manual).

Partitioning objectives are usually between the: increasing application performance, reducing network

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 11, November-2020

overhead, reducing memory constraints, saving energy, reducing friction of adoption for the programmer.

Models that are used to optimize and represent the program are Graph-based, Linear programming-based or some hybrid in-between solution.

Allocation decision are made either offline (at deployment time) or online (at runtime). The selected parts can be either statically assigned (remains unchanged during the entire life of the application) or could be dynamically modified.

Static partitioning of the code assumes that the program is being divided either during development or during first deployment. This gives a low overhead during execution but this approach is valid only when we could predict the influencing parameters accurately in advance and it is not expected to extremely vary in time.

On the other hand, dynamic partitioning allows the application deployment to adapt to the changes in the environment like bandwidth during run-time. This makes the performance of the application execution to be higher at the price of additional resource usage and latency. In dynamic offloading every time when a new optimal solution is found there will be a need for redeployment and coordination of the application between involved nodes. In case of a rapidly evolving environments, this is translated into a challenge which is to identify an adaptable partitioning of the application to offload which is still fast enough and advantageous to justify the offloading itself.

D. CLUSTERING FOR WSN AND IOT SYSTEMS

There are various clustering techniques in Wireless Sensor Networks (WSN) as well as in IoT networks. They addresses the bandwidth and connectivity for the very big and continuously growing number of devices that are attempting to connect to the Internet.

1) Clustering in WSNs: *The* clustering techniques in IoT systems refers to the topology control methods that are aiming at a more optimal usage of resources like energy, bandwidth and latency. Clustering methods achieves these motives by building up groups of devices that is to be connected such that those groups can manage collaboratively the resource usage. These methods in recent WSNs were built on a rather static approach in terms of membership of the constituting devices. The most important role in WSN clustering is the Cluster Head (CH).CH is a device in the network which is responsible for transmission scheduling. CH collects the information from all sensors in the cluster and after execution of the assembly methods forwards the pre-processed data to the gateway. Responsibilities of CH includes optimization of their work in terms of energy efficiency. That is the prolonging network lifetime through the partially switching off redundantly deployed sensors when their work is not necessary to maintain quality of service. This optimization brings two advantages:

1) Scheduling the duty cycles cuts consumptions of sensors whose work is not necessary at the given time,

2) Reducing the number of sensors trying to communicate lowers the probability of conflicts and leads to smaller latency, naturally reducing energy consumption caused by the retransmission attempts.

Forming clusters of devices can be carried out by the following steps:

1) Election of Cluster Head: Since CH is in charge of organizing the work and communication of the clusters, it will consume more energy than the others. Therefore, election should be carried out after a careful investigation. A frequent rotation between nodes is necessary for optimizing the lifespan of the network. In early systems the CH assignment is usually executed applying the following strategies:

a) Deterministic methods: They are done mostly in case of more powerful super nodes available. They have superior processing capabilities and better energy supply. They will be chosen for taking care of coordination in the network.

b) Random election: in this method the CH picked on the basis of assignment of random values which suits well for sensor fields of homogeneous devices and balanced workload.

c) Adaptive election: This refers to methods that take into account some specified parameters of consisting nodes like the residual energy or distance to base stations.

2) Cluster formation: After the selection of cluster heads, the cluster heads announce their role to neighboring nodes and the neighboring nodes will make decisions about what cluster to join. This is done according to what is the most beneficial to accomplice their own tasks and the communication range. The parameters of determination can be communication distance, physical distance, number of hops, size of the cluster and many others.

E. AI IN NETWORKING

The evolution of computer networks and communication caused a shift from static management techniques to be more flexible, robust and adaptable self-organizing technologies. These new methodologies are able to cope up with the heterogeneity and growth complexity. They aim at re-optimizing communication channels, regrouping resources in case of the radical requirement changes, or at restoring capabilities in cases of the part of the networks falls out due to malfunctions.

Due to the different possible configurations that is to be taken into consideration, it was an obvious choice to start applying Artificial Intelligence (AI) techniques to handle the high complexity in network management.

To have a common understanding of AI-managed systems, we can schematize their action with the following cycle aimed for maintaining and developing the quality of a system:

1) By monitoring, a model of the system is built up by the environment,

2) Classify the detected problems,

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 11, November-2020

3) Advocating solutions to adapt to the circumstances to achieve a better QoS.

III. REQUIREMENTS FOR A MOBILE IOT FRAMEWORK

This section is to describe about some of the requirements linked to the IoT applications and their deployment.

A. General IoT Requirements

The main challenges in IoT was always the handling of the considerable amount of data and to define services on top of that. So, IoT needs a common high-performance network with a common architectural base.

As the range of possibilities of IoT increases, the network becomes more heterogeneous and complex in shape.

In some cases, a large part of sensors time is spent in sleep mode to save energy and they cannot communicate during these periods.

However, the new smart devices produce huge amounts of real-time streaming data, and thus generating a need for effective techniques. These techniques are needed to transmit and process the data streams and to gain insights and the actionable information from real-world and measurements observations.

In contrast, some of the sensors may have to apply ad hoc communication patterns, like if they are designed to communicate only if certain rules are triggered.

In addition, these communicating devices operating with different networking standards may experience irregular connectivity with each other, some may have limited transmission range and many of them will have resource constraints. These characteristics opens up many networking challenges that cannot be solve by the traditional routing protocols.

It is important to state that the existence of IoT devices are justified only if there are applications exploiting their abilities. To stimulate the usage of their products, manufacturers have started creating software ecosystems that will enable third-parties to develop applications for their devices. Since such applications are developed by third parties, their easy integration into the platforms is a significant challenge.

Since IoT applications often do not provide complex UI, their requirements should be modelled
and resolved before their integration in the platform. The descriptions of applications and environment specifics should be built and matched in advance.

Due to the limitation of the computing resources of some IoT devices to run the application with a reasonable performance, connecting to additional computational capabilities like Edge
servers can become necessity.

All these fluctuating and heterogeneous requirements have to be satisfied through dynamic routing and coordination, transmission, a good number of access and service provisioning mechanism and by Service Provisioning Management and orchestration.

In the recent years many works have been done to resolve this issue through various architectures and orchestrator definitions. In the following section the two most similar to this is reviewed.

B. RELATED WORKS ON IOT INFRASTRUCTURES

Applications in IoT domain needs to manage and integrate great amounts of heterogeneous devices. For such a task, the IoT software ecosystems may have to use an architecture that exploits semantically enriched applications. The disadvantage of this approach is that it will result in more complex descriptions of applications and will be an increased burden for the developer.

It is necessary to have efficient task management to support IoT applications. In [3] the authors proposed a method to periodically distribute incoming tasks to increase the number of performed ones while still satisfying the QoS requirements of the tasks. The approach seemed to have better performance if the number of input tasks is large or the data size of input tasks is large, or if the connectivity of the edge network is very high. But as the author states that it needs to consider different level of prioritized inputs.

In [4] the author presents their vision and initial design efforts towards a distributed IoT orchestration architecture, but the working implementation is not presented.

The authors concentrated as the first challenge on the locality and workload aware computation partitioning between the midway servers and edge gateways and between the edge gateway and edge clients running on the corresponding edge device. They mentions the need of an intelligent task partitioning mechanism in order to enable real-time services provisioning with high scalability, but there is no proposal.

They advocated that their orchestrator should allow intelligent partition of the real time IoT computing task into an optimal coordination for server-side and IoT object processing. This will make it possible to scale it in real time when the objects are moving. It also enables the system to continuously take into consideration that the changing resource availability and workload at the distribution of computation.

The author emphasizes the need for resource-aware allocation model that could dynamically schedule and allocate resources and also a workload-aware resource scheduling of multiple services which ensures that the tasks run at the same time while taking into account of the object side processing workloads. Finally, the resource-aware selection of execution and computation environments for the collection of IoT service provisioning requests are brought up. The authors also refers to the decision on what and how to offload from edge devices to a network of IoT servers or to a cloud data centre.

In [5], the authors explored the requirements for a IoT platform through semi-structured interviews with employees having different roles in software architecture, engineering and management at their industrial partner. We could encapsulate their findings in 3 main models:

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 11, November-2020

1) Contextual Variability: In this the software should be able to adapt to the context of the device. The context which is to be modelled consists of connected devices like sensors and actuators, sensor readings like chemicals in water or features of the installation like location and temperature.
2) Modelling the Functionality of Applications in order to enable the more efficient and goal-driven UI
3) Model of the Deployment Architecture in order to express the system and contextual constraints of the applications.

All the approaches still does not consider the need for a specific modelling of the movement in order to forecast and to predict the future network topology and to answer the situations in which a fast reaction to the context changes are needed. The semantic description of the devices will be also useful for such ending.

IV. BUILDING A MOBILE IOT FRAMEWORK

From the analysis of the previous chapter, and from the background knowledge on computation offloading and application partitioning, the framework designed for offloading arbitrary tasks of IoT devices in the 5G Edge Computing environment, should implement the following five parts:

- Discovery and modelling
- Planning and optimization
- Execution
- Monitoring and performance maintenance
- Learning and predicting

a) Discovery: The first step for finding an optimal way for executing a task is profiling the context and the task itself. Profiling of a given task are carried out in many ways. The main idea of this phase is to discover the locations of the code at which the execution can be distributed or parallelized. Having an enhanced description or abstraction of the resources available, it could help with the integration from different vendors hardware and software and with fastening this profiling. An offline static analysis will help to separate monolithic applications into candidates for tasks. The scope of discovery is to build up a representation or model of the available network slice for which the deployment of the subtasks is reasonable. For this analysis we would have to gain information about the costs of transmission, the capabilities, available resources, the capabilities and current workload of the reachable nodes in the network. A combined idea of static invariable knowledge and dynamic collection of this data through simulation and estimation models is needed. Based on previous works [6] it has to be believed that a graph representation of the task connections and the cost associated in different available nodes would be more performance oriented and will be less resource intensive than a linear programming model.

b) Planning & Optimization: The results of the Discovery phase are applied to elaborate a sort of place and-route graph, as a plan to deploy our subtasks in the available network. Creation of deployment plan is very much similar to the *path computation and function placing* problem. Well known tasks must be solved by the network manager in NFV/SDN settings. As described in the previous section further restrictions to the slice of the network to consider may be posed by the application of clusters.

c) Execution: In this phase first, we have to migrate the codes of subtasks to their execution location which is done according to the execution plan built up in the previous phase. After all parts are placed on their dedicated location, the system could finally start the required computations. The challenge faced at this point by the offloading framework will be the orchestration of the collaboration of the resulting micro-services, the efficient scheduling and seamless data transmission between the nodes. This issue will be definitely addressed by using Clustering techniques. Even though there isn't a clear way for taking into considerations the mobility of the whole
system, clustering will help to reduce the managing complexity and would also improve the energy saving and the communication.

d) Monitoring: The whole infrastructure won't be dedicated to a single software of interest. The other services will be going up and down with time. This makes it necessary to monitor the performance of our system and to execute reorganization of the partitioning of some pre deployed applications for ensuring the better global performance.

e) Learning and Predicting: For the requirement of self-adaptation and self-configuration, there will be requirement for the framework to have an ability to estimate and predict the context and for the task requirements. The AI technologies will come to help in various fashions for this, as explained previously.

A. 5G EDGE FOR IOT

The systems that are made of resource constrained devices like the IoT sensors can count on offloading to handle a particularly computation intensive task. Edge Computing supports the IoT devices with a cloud closer to the edge of the network allowing some of their task to be offloaded.

The Edge servers could be part of the IoT network where the orchestration and management of the IoT devices is performed. If the Edge Computing servers are added alone then it will complicate the network infrastructure and it will not solve latency related issues completely. The Edge servers should be very flexible and should be easy to reconfigure benefiting from software defined networks and network function virtualization. The need of integrating the Edge Computing paradigm with a 5G architecture is justified by these observations. By this way, we could have high computational abilities and may get almost real time response of the system, regardless of the physical position and the capability of the involved IoT objects.

The concept of Edge Computing and 5G have been already suggested as a solution to IoT.

Such infrastructure allows the capabilities of machine to machine (M2M) communication to be inherited by IoT devices, this is a further justification for implementing IoT on top of 5G.

The classic M2M communication lacks trustworthy communication channel and the bandwidth is limited. There will be a strong dependence on the response from the server. As explained by [7], Smart Parking, Vehicle to Everything

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 11, November-2020

Communication are some of the domain areas where the IoT can take advantage of a flexible, real time and secure M2M communication. These results are achievable if supported by Edge servers and by a 5G infrastructure.

B. CLUSTERING FOR IOT

The strategies of clustering described in Chapter II are based on methods developed for WSNs; therefore the majority of those is assumed that sensors in the networks are homogeneous. The upcoming concepts IoT systems differs from this type of sensors-fields and it is aiming to serve the big variety of devices that could continuously change their position. There is a necessity of an updated resource optimizing process which will be able to adapt to this dynamically evolving circumstance for building clusters in this new environment. The backhaul issues in IoT that we face while using 3rd Generation Partnership Project (3GPP) standard communication can be grouped as follows:

1) Energy efficiency is still a problem, in static scenarios there are efficient methods to maintain QoS by shifting focus from longevity of individual sensors to the question of keeping the coverage of the network.
2) Management hierarchy: IoT has a heterogeneity of devices and subsequently a number of different services that those require in the network.
3) Data processing: The exploding amount of data generated by the ever-increasing number of gadgets is treated as one of the most valuable assets in the time of Big Data revolution. But the tremendous number of records poses a great challenge for the 3GPP networks as it tends to overload the links. In general, the transmission of entire datasets is unnecessary and highly redundant. This generated data makes it inevitable to screen and assemble before sending on to base stations.

1) CHALLANGES IN CLUSTERING FOR 5G AND IOT:

Compared to WSNs, the application of clustering techniques on IoT over 5G would raise some additional challenges. The main difference with WSNs systems is the vast heterogeneity of devices. In IoT networks huge number of transmit-only devices are present along with those super sensors, which are more powerful nodes that will be in charge of collecting and transmitting data and sometimes some pre-processing due to their computation capabilities or even organizing tasks. The super sensors will be used for the coordination of work of the cluster as Cluster Head (CH). Due to over-the-top applications, the task of CH become more comprehensive and complex in IoT systems.

Another point to be considered is the implication cost of transmission. One part of these costs is still coming from the energy costs as in WSNs but since 5G mobile network architectures is also involved we need to keep a strict control over the usage of the LTE infrastructure.

It is an important issue about how to exploit the presence of intelligent components of the core network through dynamically organizing the routing of the requests of IoT devices based on the congestion level of a given channel. In basic scenario, if it is available the most obvious choice is using wireless for transmission, but if an application has specific bandwidth requirements, we must enable the system to use assistance of LTE to provide the needed quality of experience.

While clustering in 5G network, the needs of a given user/device should be taken into consideration and based on that result of profiling clustering we should consider how to improve the user utility. The grouping together devices with similar usage could enhance the processing of transmitted data at the edge as the CH would reduce redundancies before transmitting.

The most important and relevant issue in the case of the future IoT networks is the handling of the connectivity problem of the fast-moving objects which complies the possibly high communication requirements. Previous works proved that it could be very advantageous having super sensors being able to change their locations. But the challenging question of how to organize those clusters of moving objects with the super nodes that change their position independently that is not driven by maintenance of QoS and QoE.

C. AI IN 5G

5G cellular networks surrounds a good number of access, transmission and service provisioning mechanisms. The new technologies covers topics from Radio Resource Management (RRM) through Mobility Management (MM) and Service Provisioning Management (SPM) to orchestration techniques.

The primary motives behind using AI related technologies in 5G infrastructures is that to enable the network to intelligently adjust its configurations when the requirements or parameters of the environment change.

The new 5G network will be able to provide more efficient solutions for RRM, MM, SPM, management and orchestration (MANO), making the dedicated purpose of networks no longer necessary for supporting the dynamic reconfigurations of networks as concept of NS.

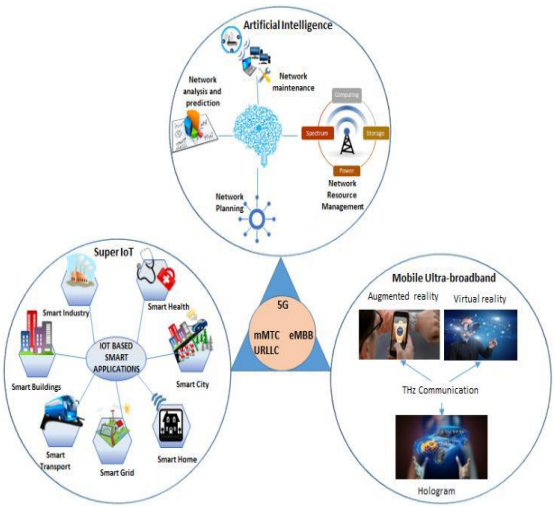

Fig 3: Architectural Scenario Where 5G Meets Artificial Intelligence

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 11, November-2020

When compared to a typical 4G node, the number of configurable parameters is expected to increase up to 1500-2000. To cope up with the new demands for self-organizing features like self-configuration, self-optimization, and self-healing it is critical that the new architecture to enhance intelligence should enable it.

Service types like eMBB, URLLC, mMTC are defined in the context of 5G are static. But as the new type of services arises continuously and patterns in existing services evolve, the system should be able to recognize these services.

The SDN architecture on 5G, with its centralized logic, provides the possibility to completely reorganize routing and network function placement. Keeping track of the state of several network devices and updating policies is very difficult when the increasingly sophisticated policies are implemented and only low-level configuration commands are available on the networking hardware. To reconfigure a network like that with the currently available tools, repeated manual interventions will be needed. In this scenario, the development of a management framework on top of 5G and SDN makes many challenges.

Such challenges can be summed up as the need for self-organization of the network which involves issues of monitoring environmental changes, learning uncertainties, elaborating and setting up new configuration of the network to maintain efficiency.

Various multidisciplinary techniques like machine learning, optimization theory, and metaheuristics can be grouped under the term Artificial Intelligence.

In [8] the authors described a possible AI-supported cellular network architecture, where the AI-controller is deployed on top of the Open Network Operating System (ONOS), or an independent network entity. The controller would communicate with Radio Access Network (RAN), Core Network (CN) and SDN controllers, and access service level agreements, information on connected user devices from SDN controllers, and collects traffic information from RAN. The AI-controller can consist of four logical modules, corresponding the MAPE (Monitor, Analyze, Plan, Execute) loop [9]:

1) Sensing module collects all the relevant information about the state of the network,
2) Mining attempts are done to discover patterns in the collected data,
3) Prediction makes predictions about the future state of the system,
4) Reasoning modules adjust parameters of the network to achieve a better performance.

An example of the application of AI controller relates to mobility issues. Sensing module tracks location of User Equipment (UE). Then using functionality of Prediction module makes predictions about the future whereabouts of the user based on the patterns of mobility developed in the mining phases. At last, according to decisions of Reasoning module, which asks for location record updates and prepares the infrastructure to handover resources to serve the predicted requests of UE.

Another possible application of AI concerned with implementation of 5G networks, is shown by the CogNet[10] project. It is an architecture for an autonomic self-managing network extending Network Function Visualization management with Machine Learning-based decision-making mechanism. The reason behind the deployment of a more adaptive controlling mechanism next to base NFV functionalities is the pursuit to reduce the costs of the system while keeping QoS at a comparatively high level.

The architecture applies the MAPE model where functionalities are implemented by the three building blocks which are data collection and storage, Cognitive Smart Engine and Policy manager [11]. The state of the network collecting records on the state of components, resource consumption of clients, and other relevant events are continuously monitored using agents. The records are then forwarded to Cognitive Smart Engine (CSE), the intelligence of the system resides in it. The main role of CSE is to process that information to decide whether and how to allocate resources, to optimize network capacity, to identify performance issues, and to secure the network. The CSE supports several Machine Learning modules to provide these services.

The engine selects the relevant parts of data and delivers it to the real-time time and batch processing engines. The engines store the result of the incoming records and based on that it makes the conclusions or predictions about the state of the system.

If necessary, the Policy Manager generates new control policies for MANO components according to the output of CSE. Control policies defines some rules on how to adjust the network configuration if they match the defined condition.

V. CONCLUSION AND FUTURE SCOPE

The work addresses the need for IoT to move towards more scalable, autonomous, connected and location independent infrastructures.

The current possibilities are defined about the development of IoT services for the network and defining and collecting the requirements for a fully trustworthy combination of IoT technologies and 5G that allows to exploit the most of both.

The relevant background knowledge on previous efforts on computation offloading and application partitioning, with state-of-the-art challenges for IoT have been brought together. The possible solutions to those through complementary technologies and paradigms such as 5G Edge Computing and the integration of Clustering and AI methods.

5G Edge architecture responds to the needs of IoT networks with a fast-reliable communication with less overhead, facilitating the computation offloading by reducing to minimum link costs and allows the better management of the whole infrastructure adding decentralization and thereby reducing the need to communicate with Central Data centres.

The clustering techniques helps to reduce the communication loads at the edge of the network which will help to save energy and simplify the network management in a *divide and conquer* fashion.

The AI based technologies enhances the infrastructures ability to adapt to the continuously changing requirements and reorganizes itself whenever necessary.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 11, November-2020

5G-enabled IoT is expected not only to enable technological growth but also to bring jobs to about 22 million people.5G could enable IoT to improve security and public safety. It will improve healthcare, bring smart Cities Intelligent traffic control and self-driving cars to reality.

REFERENCES

[1] Peter Kiss, Anna Reale, Charles Jose Ferrari, Zoltan Istenes" Deployment of IoT applications on 5G Edge"
[2] Suresh Borkar;Himangi Pande "Application of 5G next generation network to Internet of Things "
[3] Y. Song, S. S. Yau, R. Yu, X. Zhang and G. Xue," An Approach to QoS-based Task Distribution in Edge Computing Networks for IoT Applications," 2017 IEEE International Conference on Edge Computing (EDGE), Honolulu, HI, 2017, pp. 32-39.
[4] Applications 48 (2015): 99-117. E. Yigitoglu, L. Liu, M. Looper and C. Pu," Distributed Orchestration in Large-Scale IoT Systems," 2017 IEEE International Congress on Internet of Things (ICIOT), Honolulu, HI, 2017, pp. 58-65.
[5] M. Tomlein and K. Grnbk," Semantic Model of Variability and Capabilities of IoT Applications for Embedded Software Ecosystems," 2016 13th Working IEEE/IFIP Conference on Software Architecture (WICSA), Venice, 2016, pp. 247-252.
[6] J. Liu, E. Ahmed, M. Shiraz, A. Gani, R. Buyya and A. Qureshi, 2015. Application partitioning algorithms in mobile cloud computing: Taxonomy, review and future directions. Journal of Network and Computer Applications, 48, pp.99-117. 30 -R. Li et al.," Intelligent 5G: When Cellular Networks Meet Artificial Intelligence," in IEEE Wireless Communications, vol. PP, no. 99, pp. 2-10.
[7] S. H. Shah and I. Yaqoob," A survey: Internet of Things (IOT) technologies, applications and challenges," 2016 IEEE Smart Energy Grid Engineering (SEGE), Oshawa, ON, 2016, pp. 381-385.
[8] R. Li et al.," Intelligent 5G: When Cellular Networks Meet Artificial Intelligence," in IEEE Wireless Communications, vol. PP, no. 99, pp. 2-10.
[9] J. O. Kephart and D. M. Chess," The vision of autonomic computing," in Computer, vol. 36, no. 1, pp. 41-50, Jan 2003.
[10] I. G. Ben Yahia, J. Bendriss, A. Samba and P. Dooze," CogNitive 5G networks: Comprehensive operator use cases with machine learning for management operations," 2017 20th Conference on Innovations in Clouds, Internet and Networks (ICIN), Paris, 2017, pp. 252-259.
[11] L. Xu et al.," CogNet: A network management architecture featuring cognitive capabilities," 2016 European Conference on Networks and Communications (EuCNC), Athens, 2016, pp. 325-329.
[12] Kinza Shafique;Bilal A. Khawaja;Farah Sabir;Sameer Qazi;Muhammad Mustaqim," Internet of Things (IoT) for Next-Generation Smart Systems: A Review of Current Challenges, Future Trends and Prospects for Emerging 5G-IoT Scenarios"