**A Review on Big Data Analysis and Internet of Things**

Juliya P Jerom, Felix Rohan Abraham

Department of Computer Applications

Saintgits College of Applied Sciences

Pathamuttom, Kottayam-686532

juliyapjerome@gmail.com felixrohan2002@gmail.com

**ABSTRACT**

Internet is continuously evolving and changing. IoT connects billions of devices to the Internet. These devices are generating enormous amount of data. IoT can be considered as the future of internet applications which involves machine to machine learning. Big Data and IoT can be viewed as two sides of a coin. With the connection between Big Data and the objects on Internet benefits of IoT can be easily reaped. IoT applications are used in a wide range of industries, including disaster management, retail management, and health care. This data must be transmitted and processed, which is a difficult operation. This paper aims at reviewing the role of big data in IoT through discussion of its protocols and architectural structure. Methods for testing these protocols and security considerations are presented. Multiple IoT applications and future research directions are also discussed.

**KEYWORDS:**

Internet of Things (IoT),Big Data, Machine to Machine Learning.

**INTRODUCTION**

Nowadays, internet apps are strongly related to daily life. Several apps, from e-mails to e-learning, are popular and have sped up otherwise difficult processes. Technology is being improved so that it can easily fit into everyday life. The ability to install a variety of applications that can simplify and facilitate complex operations across a range of industries has been made possible by the digital world. Everyone can now quickly obtain the information they need thanks to the Internet, but its use is evolving as more network resources become available. Large amounts of Internet traffic will be produced by non-PC machines as they communicate with other machines. Non-PC machines will generate large volume of the Internet traffic to transmit data to other machines. The Internet of Things is made up of the complete network of non-PC communication, such as signals between M2M and Machine to User (IoT). IoT is an expanded use of the network resources currently available, extending communication from users to users to users to things and things to things. As a result, numerous computing and sensor devices of all kinds have been deployed.

IoT may be summed up as a global network of individually accessible connected devices that can communicate with one another utilising already-existing or newly developed technologies in order to be controlled by conscious choices. The term IoT was coined by Kevin Ashton in 1999. These devices are becoming smarter as a result of their ability to directly communicate with one another and interpret locally generated data for intelligent decision-making based on predefined algorithms.

**IMPORTANCE OF BIG DATA IN IOT**

IoT generates large data in a variety of fields due to improving data processing capabilities, abundant data storage resources, and ultra-fast Internet. Sensor devices are typically used to gather data, and these devices send that data to a centralised server. The data is likewise delivered back to the devices in a similar way. Generally, big data is characterized by 3 Vs: Volume, Variety and Velocity. IoT not only increases the volume of data traffic but also involves diverse data from heterogeneous devices. Furthermore, due to the involvement of big data in IoT, high transmission rate and proper analysis of data for correct functioning of devices remain a challenging task. Intel in one of its report states that the IoT devices are generating large amount of unstructured big data which will be useless if there is no algorithm to properly analyse it. There are three important steps that are involved in big data analysis: storage, processing and accurate result generation. Traditionally, the storage of data is done through Extract, Load and Transform technique. The ELT technique's lack of adaptability for new data sources in terms of data storage and processing makes it unsuitable for dynamic IoT networks.

Big data management for IoT is a crucial field that entails structured data analysis. There are various tools that are designed by different organizations to develop big data management system. One of them is Apache Hadoop, which use the Map Reduce paradigm to process massive data from various devices. To enable processing to be carried out on a variety of devices using a parallel processing algorithm, Hadoop divides the data sets into numerous sets. Moreover, Hadoop enables hundreds of processors to offer local storage and processing capacity to these sets, enabling big data analysis for the Internet of Things. As a result, all local IoT devices can process massive amounts of data by pooling resources. These devices need to transmit data through different modes of communication which will be discussed in the next section.

**MODES OF COMMUNICATION**

As was previously said, the Internet of Things (IoT) is a network of addressable objects that require a communication channel to coordinate with one another and send data. Moreover, IoT is a division of Cyber Physical System (CPS) and has its own well-established protocol for communication. We'll talk about various protocols that are currently in use in the CPS domain and how they may be used to send large amounts of data between IoT devices.

Because big data is involved in IoT, the primary objective is to process data locally rather than through the cloud. Hence, the Localized Cooperative Access Stabilization (LCAS) algorithm, which is made for devices to communicate locally, is one of the potential algorithms for localised processing. In this algorithm the nodes, devices, are not only aware of the main server state but are also aware about the state and address of the neighbour nodes. By employing this approach, the nodes can communicate with one another locally to transfer large amounts of data without contacting the main server, which helps conserve network traffic. LCAS can be implemented in IoT domain in multiple ways-exact, adhoc, hierarchy and hierarchy +adhoc schemes.

The exact scheme allows the devices to operate as a bridging node for communication between other devices in addition to sending and receiving data locally via LCAS. The closest nodes to the primary local server must regularly function as a bridge for data transfer, which makes this scheme the least energy-efficient when compared to others even though it takes full use of the local resources available there. All of the sensors in an ad-hoc-based system can function as nodes for massive data transmission, but the nodes also know the optimum route for moving data from the source to the destination node. In order to calculate the shortest path all the nodes frequently broadcast a short message and then from the response time of each node it calculates the shortest available active path. This scheme not only utilizes all the available resources but also it is more efficient in terms of data transmission rate than exact scheme. In contrast to these two schemes, the devices are split into upper and lower layers in a hierarchy scheme. Lower layer devices only function as senders or receivers of their own data, while upper layer devices serve as nodes that transmit data for lower layer devices in addition to transmitting their own data. On the basis results discussed, this scheme has higher data transmission rate as compared to ad-hoc scheme. Furthermore, the hierarchy + ad-hoc scheme is similar to hierarchy scheme but the devices in lower layer can also communicate with their immediate neighbours locally which makes it the most efficient scheme among all others in terms of data rate and energy consumption. Although all of these approaches allow for continuous data transmission across all devices, there are some IoT scenarios where this is undesirable. Threshold sensitive Energy Efficient Protocol (TEEP) and Threshold sensitive Stable Election Protocol (TSEP) are created for devices that can sense data at all times but can only send data when its value is greater than the predetermined threshold in order to address this issue. Although these protocols are energy-efficient, if the threshold is not met, they will never transfer data. The devices should be able to automatically communicate the data to the primary local server or to other nodes on a regular basis in order to employ TEEP and TSEP for IoT to send huge data. This enables prompt and accurate data collection from the devices for utilisation. For IoT communication utilising the LCAS algorithm, the suggested modified TEEP and TSEP protocols are the best available solutions. Once different schemes for IoT to transmit big data are analysed then an important step towards the development of communication between different nodes is to select the mode of communication. Since the IoT generates a vast amount of data, a standardised protocol that may maximise the utilisation of resources between nodes should be employed. An intriguing remedy for the IoT's bandwidth and spectrum inefficiencies is Cognitive Radio Technology (CRT).

Mobility of IoT is another important factor that plays an important role in the decision of selection of mode of communication. IoT devices range from being static to being very dynamic in networks, in addition to having heterogeneous characteristics. Thus, selecting correct mode of communication for IoT involves multiple factors and efficient data communication rate can be achieved through combining these discussed protocols. The next section will discuss in detail the layered architecture of IoT in order to use these protocols for communication.

**LAYERS OF IOT**

IoT devices produce a huge amount of data, hence the Internet's current architecture needs to be modified to create an effective communication protocol. There isn't one globally accepted IoT design that everyone agrees upon. Various researchers have presented several architectural designs. Chen Min discusses three-layer architecture for IoT while Suo et al. divide the IoT architecture into four layers. The ability to process data at each layer due to the architecture's split into layers reduces the need for big data processing at the top layer. The six-layered Zhang et al. design is depicted in Figure 1. In light of what follows, this design is appropriate for the Internet of Things.

The first layer that provides the unique identification to IoT is the coding layer. To maximise data rate, the routing and communication algorithm are chosen using the data collected from this layer. This layer is appropriate to detect the effective data rate requirements of each device because the Internet of Things network is heterogeneous. The physical layer of the devices that gather and receive data is, in essence, the perception layer. It is suggested that the perception layer for the Internet of Things be upgraded by adding data processing capability so that processed information can be sent to higher layers. This way the quantity of big data can be decreased as only processed information is transmitted to destination. Also, the primary function of the network layer is to provide as a link between the middleware layer and the perception layer. The majority of big data analysis is now carried out at the middleware layer, which slows down the IoT network. Hence, this layer needs a lot of computing power. The application layer then generates responses based on the processed data when the middleware layer has finished processing the data. The business layer, which oversees a device's applications and services, is also crucial to the network's overall efficacy. As a result, the layer architecture, which is the core concept of the Internet of Things, is crucial in reducing the amount of data transfer to the Internet.

The next crucial step is to test the designed protocols, which will be covered in the following section, after they have been created for communication while taking into account the layer architecture.



Fig.Layers of IoT

**PROTOCOL TESTING**

IoT testing comprises doing QA tests to examine the performance, security, and functionality of IoT devices. It is crucial to ensure that your IoT devices can transmit sensitive information wirelessly before going to market because every IoT device sends and receives data over the Internet. To find flaws before they reach customers, many IoT organisations rely on IoT automation, penetration, and performance testing solutions. IoT testing seeks to make sure that IoT devices adhere to given specifications and perform as anticipated. The testing of these suggested communication protocols is necessary to see whether they are appropriate for the IoT stream, as was previously discussed.

An open-source network planner built on Java called Net2Plan can be used to evaluate IoT built-in protocols. Also, this tool enables users to develop unique algorithms.

Users may easily monitor the condition and status of the complete Internet structure, including delays, route, etc., thanks to Net2Plan's network representation service. Network designers can create protocols to maximise the transfer of huge data by observing the effects of data transmission rate on the Internet.

Another platform for testing protocols is EuWIn. Three well-known IoT communication protocols—Software Defined Wireless Networking (SDWN), ZigBee, and IPv6—were examined by Buratti et al. The collected results suggest that SDWN performs better for static and semi-static devices while for highly mobile devices data transmission rates for ZigBee are higher.

SensorHUB is another helpful IoT protocol testing platform. This platform is helpful since it enables the ability to examine data between different nodes in addition to providing a space to monitor IoT-related communication. Therefore, this platform can be used for the development of protocols while focusing on big data analysis

**APPLICATIONS OF IOT**

When Internet-connected IoT devices are linked together, it is possible to create apps that allow for the collection and use of huge data from several devices. Since these devices generate real time data therefore, they are an active source for big data generation. Also, the extremely varied nature of devices enables the creation of IoT applications in a variety of fields by developers. Currently, many developers are actively working in the domain of health 328 care system, smart building and connected vehicles to design IoT network for the development of these areas.

IoT in agriculture allows monitoring and management of microclimate conditions for indoor planting a reality, which in turn boosts yield. IoT-enabled devices can monitor soil moisture, nutrients, and meteorological information to better manage irrigation and fertiliser systems for outdoor planting. For instance, this prevents resource waste if sprinkler systems only release water when necessary.

Wearable IoT devices are used in the healthcare system to monitor patients' health while they are at home, eliminating the need for hospital visits while still giving up-to-the-second real-time information that potentially save lives. Smart beds reduce the wait time for available space in hospitals by keeping the personnel informed of availability. The installation of IoT sensors can increase reliability and decrease malfunctions, which can be the difference between life and death. IoT makes providing care for elderly people much more comfortable. Sensors can identify whether a patient has fallen or is having a heart attack in addition to the previously described real-time home monitoring.

By creating methods for intelligent transportation systems, the automotive sector has also demonstrated improvement in the design of IoT systems for linked automobiles. The development of linked vehicles has made it feasible for automobiles to communicate with one another, with the Internet, with connected sensors, and with road infrastructure. This allows for the monitoring of both traffic conditions to prevent congestion as well as road infrastructure to monitor and enhance vehicle safety. As the number of sensors in this field will increase the amount of data generated by them will also increase exponentially. IoT has the ability to actively contribute to the upgrading of numerous domains' infrastructure and can support more effective resource management. Due to the addition of sensors in many applications, a significant amount of research is required for the handling of that data.

**CHALLENGES IN IOT**

The Internet of Things (IoT) refers to the network of interconnected devices that can communicate with each other and exchange data. While IoT has tremendous potential to transform various industries, there are several challenges associated with its implementation and adoption Some of these challenges include:

Security: One of the biggest challenges in IoT is ensuring the security and privacy of the devices and data. IoT devices are often vulnerable to cyber-attacks, and securing them can be a complex and ongoing process.

Interoperability: IoT devices are often developed by different manufacturers, and they may use different communication protocols and data formats, which can make it difficult for them to work together seamlessly.

Scalability: As the number of IoT devices and data generated by them continues to grow, managing and analysing the data can become a significant challenge.

Power Consumption: IoT devices often run on batteries or low-power sources, and their constant communication with other devices can drain their power quickly.

Cost: The cost of IoT devices and the infrastructure required to support them can be a significant barrier to adoption, especially for small and medium-sized businesses.

Data Privacy: IoT devices often collect and transmit sensitive data, which raises concerns about data privacy and ownership.

Integration: Integrating IoT with existing systems and processes can be a significant challenge, and it often requires significant changes to the way organizations operate.

Standards: There are currently no standardized protocols or guidelines for IoT development and implementation, which can lead to compatibility issues and hinder the adoption of the technology.

Addressing these challenges will require a collaborative effort from manufacturers, policymakers, and other stakeholders to ensure that IoT is developed and deployed in a safe, secure, and effective manner.

**CONCLUSION**

The Internet of Things (IoT) and Big Data are two interconnected technologies that are transforming the way we live and work. IoT refers to the network of interconnected physical devices that collect and exchange data, while Big Data refers to the large volume of data generated by these devices and other sources. The combination of IoT and Big Data has created new opportunities for businesses to gather and analyse data, gain insights, and improve decision-making. Big Data technologies such as data analytics, machine learning, and artificial intelligence can then be used to process and analyse this data, generating insights and driving innovation.

However, the adoption of IoT and Big Data also presents a number of challenges. The sheer volume of data generated by IoT devices can be overwhelming, and businesses must have the infrastructure and expertise to manage and analyse this data effectively. There are also concerns around data privacy and security, as IoT devices may collect sensitive information that needs to be protected.

In conclusion, the combination of IoT and Big Data has the potential to transform industries and improve our lives in many ways. However, businesses must be prepared to address the challenges and risks associated with these technologies in order to realize their full potential.