An "Intelli-Fog" Approach To Managing Real Time Actionable Data In IoT Applications

Ekpe Okorafor[1] **and Mubarak Adetunji Ojewale**[2]
[1, 2] Department of Computer Science, African University of Science and Technology, Abuja, Nigeria

Abstract - *The Internet of Things (IoT) has generated a large amount of research interest across a wide variety of technical areas. These include research interests in the physical devices, communications between the devices, and relationships between them. One of the effects of ubiquitous sensors networked together into large ecosystems has been an enormous flow of data supporting a wide variety of applications. Data stream processing frameworks enable us get real time insights from these enormous data generated by IoT applications. Fog Computing has been proposed and implemented for IoT applications as it brings decision making closer to the devices and reduces action latency among many other advantages. The low computation power, typical of edge devices, also makes them unsuitable for complex stream analytics. In this work, we propose a new "Intelli-Fog" approach to IoT data management by leveraging mined historical intelligence from Cloud Big Data platforms and combining it with real time actionable events from IoT devices at the edge of the network. IoT devices export data to Big data platforms at intervals and the platforms in turn communicate information needed for real time intelligent decision making. We also present use case application scenarios. This approach makes edge devices more intelligent in decision making without increasing action latency.*

Keywords: IoT; Fog Computing; Real Time events

1 Introduction

Advances in sensor technology, communication capabilities and data analytics have resulted in a new world of novel opportunities. With improved technology, such as nanotechnology, manufacturers can now make sensors which are not only very small to fit into almost anything, but also more intelligent. These sensors can now pass their sensing data effectively and in real time due to improvements in communication protocols between devices. In the era of Big Data and the Cloud, there are now, also, emerging tools for storing and processing the increasing amounts of data. These phenomena combined with the need to gain insight from the data, has made the Internet of Things (IoT) a topic of interest among researchers in recent years. Simply put, IoT is the ability of people "Things" to connect with anything, anywhere and at any time using any communication medium. "Things" here means connected devices of any form. It is estimated that by 2020, there will be 50 to 100 billion devices connected to the Internet [1]. These devices will equally generate an incredible amount of massively heterogeneous data. The data, due to size, rate at which they are generated and heterogeneity is referred to as "Big data". Big data can be defined with 3 characteristics known as the 3Vs; volume, variety, and velocity or sometimes 5Vs including Value and Veracity [2], [3]. Big data, if well managed, can give us invaluable insights into the behavior of people and "things"; an insight that can have a wide range of applications.

The potentials for incorporating insights from IoT data into aspects of our daily lives are becoming a reality at a very fast pace. The acceptability and trust level is also growing as people have expressed willingness to apply IoT data analytics results in important decision making domains; domains as important as stock market trading [4]. These developments inform the need for efficient approaches to manage and make use of the huge and fast moving data streams in a way that extracts real time value from it. Distributed processing frameworks such as Hadoop have been developed to manage large data but not data streams. One major limitation of distributed platforms such as Hadoop is with latency. They are still based on the traditional Store-Process-and-Forward approach which makes them unsuitable for real time processing; a contrast with the real time demands of the current and emerging application areas [5]. Store and forward approaches also will not be able to satisfy the latency requirements of IoT data because of the velocity and the unstructured nature of the data. Stream processing frameworks like Apache Storm and Samza have been introduced to solve this problem. In stream processing, data from data sources are continuously processed as they arrive and do not need to be first stored. These stream processing frameworks usually leverage Cloud and Distributed Computing. IoT data are typically in streams and suits stream processing applications. Actionable events from IoT applications, however, have strict latency requirements than Cloud stream processing frameworks can provide. So, combining mined intelligence from sensor data with real time events from IoT to make intelligent decisions in real time requires a much faster approach; an approach that takes decision making closer to the sensing devices.

To achieve this, we propose Fog-Computing to complement Cloud data management infrastructures in capturing and reacting to real-time actionable events. We propose an "Intelli-

Fog" layer in which mined intelligence from Big Data Platforms can be cached at the edge of the network and available to the devices for real-time intelligent decision making. This information (the mined intelligence) is updated at intervals regular enough to still make it relevant for correct decision making.

2 Fog Computing and IoT

One can think of Cloud computing as a manufacturing company (an application) where all customers must buy goods (access services) from the factory (cloud) alone. The Fog computing would be the company deciding to open retail outlets at locations close to the customers and the retail stores will give a level of access to the customer; if not exactly, it will be close to what they can obtain from the factory. The Fog Computing paradigm extends Cloud computing and services to the edge of the network; close to the devices. It is a hierarchical distributed platform for service delivery providing computation, communication, and storage at a layer that is much closer to the devices than the cloud. It is not also "just" a trivial extension of Cloud computing. It has some distinctive characteristics that make it more than just an extension. These characteristics per Bonomi [6], include Edge location, location awareness, and low latency, Geo-distribution, Mobility, Real time interaction among others. Location awareness, most especially, makes fog-computing more than just an extension of the Cloud.

Fog Computing is described as a highly virtual platform that provides compute, storage, and networking services between end devices and traditional Cloud Computing Data Centers, typically, but not exclusively located at the edge of network [6]. Fog computing is not meant to replace Cloud computing; rather, it is to compliment it in applications where the traditional Cloud computing may not be suitable. Examples of such applications include Geo-distributed applications, large-scale distributed control systems, and applications with very strict latency requirements [7].

The early approach to managing IoT data is for the device to send data directly to the cloud via a communication medium, in real time. Big data analytics is also done at the cloud layer and decisions, where applicable, are communicated to the devices from the cloud to the devices via the communication medium also. IoT applications are typically highly distributed and the velocity of data is high. This can cause congestion on the network and communicating actionable events and decisions can be a challenge. An actionable event can get caught in a busy channel and can be delayed with other messages. This is a major contributing factor to action latency, even when the channel is free. The distance between the cloud layer and the devices is usually too wide. Processing actionable events at a layer closer to the devices was suggested and that introduced the Fog concept to IoT data management.

The fog layer enables distributed processing at the edge of the network; very close to the devices. The Fog approach reduced latency significantly and brought processing close to the edge of the network. It is also sometimes used for dynamic real-time load balancing. Processing of data geo-sensitive data is also easier and data is aggregated before transmission to the cloud layer for further processing. However, the existing Fog approach does not provide a mechanism for making intelligent decisions in real time while taking mined intelligence from Big data analytics into consideration. Decisions are typically based on the actionable events alone and/or the real-time state of the system.

3 The Intelli-Fog Approach

The intelli-Fog Layer of the proposed approach differs from the existing fog approaches because it can make more intelligent decisions; without any significant latency increase. This layer communicates both with the IoT devices as well as the Cloud layer where data analytics is performed. Communication with devices involves taking in raw data/actionable events that are observed or reported by IoT devices. It also involves communicating decisions/actions taken back to the IoT devices. This is done in parallel with communication with the cloud layer. The communication with the Cloud layer involves a two-way mode. Aggregated data is sent to the cloud for storage and analytics and mined intelligence from the historical data are sent back to the intelli-Fog layer for real-time decision making. Other information necessary for decision making is also available at the intelli-Fog layer. The proposed Intelli-Fog approach is shown in Figure 1 below.

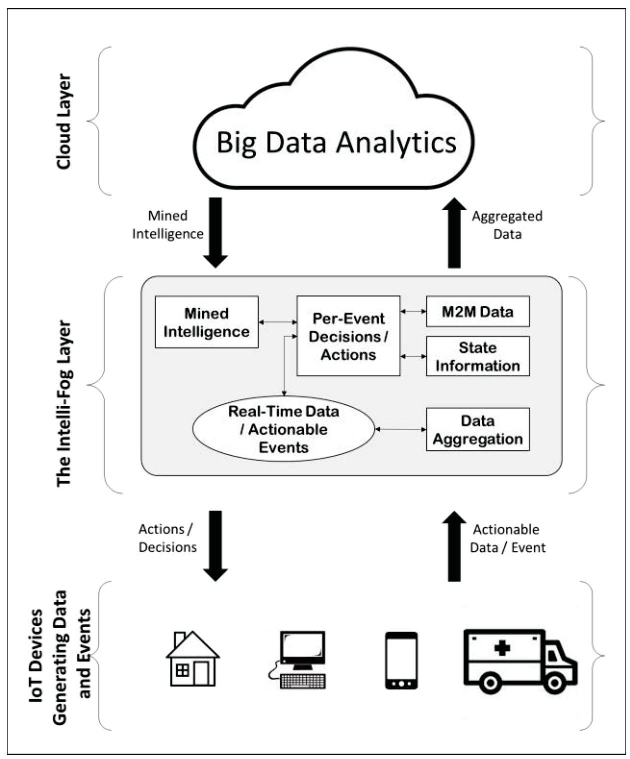

Figure 1: The Proposed Intelli-Fog Approach

4 Application Scenarios

4.1 Intelligent Patient Monitoring System

A patient monitoring machine will sound an alarm by sensing an abnormal reading on patient data; but the alarm cannot state the likely symptoms the health care officer should look out for immediately. With the proposed system, a patient monitoring system can be made more intelligent to make these decisions based on some analysis on patients' medical history as well as other available relevant medical records to predict likely complications to watch out for. An intelli-Fog layer can also go further to communicate this reading to the patient's doctor or a doctor on duty; and communicate the patient's necessary medical record too for speedy, accurate and efficient health care service.

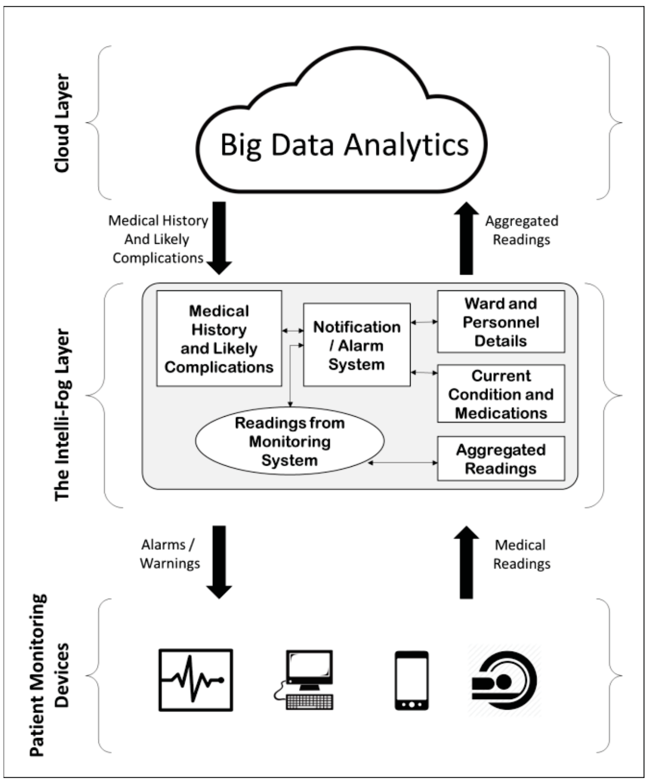

Figure 2: Intelligent Patient Monitoring System with the Intelli-Fog Approach

4.2 Intelligent Traffic Monitoring

Smart Traffic Lights serves as the intelli-Fog layer of a traffic management system. The Smart traffic light senses both vehicles and pedestrians on the road. It monitors the state of each lane and uses Machine-to-Machine (M2M) communication protocol with other Smart Traffic Lights to get the state of neighboring states. It aggregates the data and then sends to a central cloud layer for data analytics. The cloud layer sends back information on likely busy lanes including periods and suggestions on how to pass cars in each lane. This information, together with the state of other lanes is used by a Smart Traffic Light to pass cars on its lane.

5 Related Work

Some work has been done with respect to IoT data management but, in this section, we focus on the ones related to data, analytics and performance.

Khan et al. [8] proposed a cloud-based data management and analytics service for a cloud city application. They proposed an architecture that provides basic components to build necessary functionality for a cloud-based Big Data analytical service for smart city data. This architecture was implemented in both Hadoop and Apache Spark on the Bristol open data. The results show that Hadoop incurs more overhead, especially in job submission, than Spark and that Spark is more appropriate for the Bristol open data. The part of the open data analyzed is the data about quality of life measured by indicators such as crime rate, security, etc. Their work is an important contribution to the smart cities projects, a very popular application of IoT. Their work, however, does not consider real-time analytics.

Khodadadi et al. [9] proposed another cloud-based data-centric framework for development and deployment of IoT applications. The framework architecture manages data collection, filtering and load balancing from different sources, producing both structured and unstructured data. It was demonstrated with an application to compute tweet sentiments of the six biggest capital cities in Australia. This application was built on top of the Aneka Cloud Application Platform. This framework focuses on data collection and its abstraction from developers and it is an important contribution towards finding a generic approach to IoT data collection management. It, however, does not deal with data analytics.

Zhu et al. [10] proposed a framework to make disparate and incompatible datasets usable, interoperable and valuable across the enterprise. This framework, though not specifically for IoT, enables data from disparate sources to be more "homogeneous" by proposing a Common Information Model (CIM) standard for information interchange between data sources. This approach will require data sources to use the CIM adapters; a condition many IoT devices may not be willing or able to meet. This approach is not also concerned with performance as it is mainly concerned with making data more homogeneous from the source for easy analytics. It is also specifically designed for the utility industry and not a general framework for Big Data analytics.

Abu-Elkheir et al. [11], having investigated the issues of IoT data management, also introduced a model framework for IoT data management. This model provided layered stages of data management. The work proposed six layers of IoT data management including a "things layer", communication layer, source/data layer, federation layer, query layer and application layer. Their work also does not involve any specific handling of real-time processing and optimization issues. It focused mainly on how data is collected, stored and processed. The model is also focused on batch processing.

Cecchinel et al. [12] also developed an architecture support for a data collection framework in IoT. This framework handles IoT Big Data issues such as sensor heterogeneity, scalability data and even reconfiguration capability. The use was also used demonstrated on sample projects. The work, however, focuses on the data collection part and does not consider the data processing. Also, the framework handles data aggregation at the bridge to make data arrive at the middle-ware in a unified format. Their approach will work fine if the IoT data management application is location specific and data always arrive through the bridge. It may not work when data sources are widely dispersed and data come in directly through the internet without passing the bridge.

The Lambda architecture [13] was proposed by Nathan Marz. The architecture is generic for large stream processing, giving the options of real-time processing and batch processing concurrently. In the Lambda architecture, incoming data is dispatched to both the batch layer and the speed layer for processing. The batch layer manages the master dataset (an immutable, append-only set of raw data), and pre-computes the batch views. The service layer indexes batch views for low latency, ad-hoc queries and the speed layer deals with recent data and real-time analytics only. Incoming queries can be typically answered by merging the results from batch views and real-time views.

The Lambda architecture received some criticisms, however, especially in terms of its complexity [14]. Some others criticized the redundancy in implementing almost identical processes in both layers. Lambda architecture is also criticized for the one-way data flow and immutability of data. Lambda's shortcomings also include its inability to build responsive, event-oriented applications.

6 Conclusion and Future Work

In this work, we have proposed an "Intelli-Fog" layer which caches mined intelligence from Big data analytics and makes it available for real-time decision making at the edge of the Network. This makes intelligent response to actionable events faster and closer to the devices. Immediate decision making and Big data analytics can also be taking place concurrently in this new approach with just a single data entry point and a very simple architecture. In the future, we want to work on demonstrating the approach by implementing a use case example and comparing the benchmark with existing approaches

7 References

[1] C. Perera, A. Zaslavsky, P. Christen and D. Georgakopoulos, "Context Aware Computing for The Internet of Things: A Survey," Communications Surveys Tutorials, IEEE, vol. 16, no. 1, pp. 414- 454, 2013

[2] C. Eaton, D. Deroos, T. Deutsch, G. Lapis and P. Zikopoulos, "Understanding Big Data", McGraw- Hill Companies, 2012.

[3] Yuri Demchenko "Defining the Big Data Architecture Framework (BDAF)", University of Amsterdam, Amsterdam, 2013.

[4] C. Perera , R. Ranjan,L. Wang, S. Khan , and Y. Zomaya, "Big Data Privacy in the Internet of Things Era" Issue No.03 - May-June (2015 vol.17) pp: 32-39

[5] M. Shah "Big Data and the Internet of Things" Research and Technology Center - North America, Palo Alto, 2015.

[6] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli "Fog Computing and Its Role in the Internet of Things", in Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing (New York, NY: ACM), 2012.

[7] M. Pranali "Review of Implementing Fog Computing", International Journal of Research in Engineering and Technology (IJRET), Volume: 04 Issue: 06, 2015.

[8] Khan Z., Anjum A., Soomro K., and Tahir M., "Towards cloud based Big Data analytics for smart future cities" Journal of Cloud Computing: Advances, Systems and Applications, 2015.

[9] Khodadadi, F.; Calheiros, R.N.; Buyya, R. "A data-centric framework for development and deployment of Internet of Things applications in clouds" (ISSNIP) 2015 IEEE Tenth International Conference, 2015

[10] Zhu J., Baranowski J., Shen J., "A CIM-Based Framework for Utility Big Data Analytics" (2014) available at: http://www.powerinfo.us/publications/CIM_Based_Framework_for_Big_Data_Analytics.pdf

[11] Abu-Elkheir M., Mohammad Hayajneh and Najah Abu Ali, "Data Management for the Internet of Things: Design Primitives and Solution" sensors ISSN 1424-8220, 2013.

[12] Cecchinel C., Jimenez M., Mosser S., Riveill M., "An Architecture to Support the Collection of Big Data in the Internet of Things" IEEE International Workshop on Ubiquitous Mobile Cloud, 2014.

[13] http://lambda-architecture.net/

[14] Richard Hackathorn, "The BI Watch: Real-Time to Real Value," DM Review, 2004.