# Data Fusion for the Internet of Things

**Hemanth Kumar** [*], **Pratik Pimparkar** [**]

[*] Embedded Engineer, Ideal Educraft Technologies, Bangalore, India
[**] Embedded Engineer, Ideal Educraft Technologies, Bangalore, India

DOI: 10.29322/IJSRP.8.3.2018.p7541
http://dx.doi.org/10.29322/IJSRP.8.3.2018.p7541

***Abstract-*** Internet of things has been increasing at a tremendous rate since last few years. Thus, it has been an active area of research. Data from a single node can provide enough information for various conditions on the field. However, the data from the IoT is used to get an inference on a high level which can be used to take corrective actions. To take the corrective actions, the data should be mined properly which in turn depends on correct data from the sensor. Data fusion techniques are used to provide correct data from the sensor to the data mining algorithm. Data fusion in IoT is not a well-researched topic. In this design, various data fusion techniques are examined, and a hierarchical approach for data fusion in IoT is proposed. The design uses Fuzzy Kalman Filter for state estimation, and Dempster- Shafer method for decision fusion to create a dynamic context-aware system. The proposed design is also scalable for a higher number of nodes in the network which is not found in all implementations. The research also offers a design metrics which can be used for comparison of different data fusion.

***Index Terms***- Internet of Things, Data Fusion, Kalman Filter, Dempster – Shafer Theory.

## I. INTRODUCTION

The Internet of Things (IoT) generates tremendous amounts of raw data from its surrounding environment. Wireless Sensor Networks which form a part of IoT sense the environmental conditions and relay the information to the server. This information is then processed using data mining algorithms at the server side to get an overall picture of how the conditions are and then control actions are taken. Usually, the data provided by the sensor nodes is raw. Significant computation needs to be done on the data before meaningful extraction of information can take place. The computation should be performed before the data mining can take place. The data mining is the final step of information extraction. But before the data is fed for mining it needs to be in the proper format so that the predictive pattern or anomalies in the data can be found. This data processing technique which combines and integrates data from different sources is called as data fusion. With data fusion, it becomes easier to take decisions based on high-level inferences from the data extraction. Also, less time and computational capacity will be expended to get to an inference if the data is meaningful.

Data fusion is not a new concept and has been implemented in different domains like image processing for data extraction. There is a lot of research on IoT data fusion techniques which study the fusion of data in Wireless Sensor Networks. These methods use mature data fusion algorithms like Bayesian, Classical inference, Dempster – Shafer, and fuzzy logic approach to implementing it for Internet of Things (Rodger, 2012). However, research should also to go into the hierarchy of data fusion implementation. The sub-layers of architecture in which the data fusion techniques are implemented are very important. This domain is not well explored, and the research intends to put light on postulating architecture for data fusion.

In this research, the current implementations of data fusion techniques will be studied in detail. Their hierarchical structure will be explained and compared with standard parameters. Using the knowledge gained in this study, another hierarchical solution will be provided, and it will be compared with the existing techniques. This research will extend the hierarchical implementation of IoT and cover all levels of data abstraction i.e. from a sensor to server.

## II. RELATED WORKS

Data Fusion has been implemented in different domains for sensors from heterogeneous backgrounds. It is defined by Castanedo (2013) as “Combination of multiple sources to obtain information which is less expensive and has higher quality or more relevant information.” The following section describes the existing methods for multi-sensor data fusion model. This section discusses the methods in detail and shows the gap that the research intends to fill.

Wide research has been done in Multi-Sensor Data Fusion which has application in various sectors. There are different data fusion techniques implemented by different researchers. There are statistical methods employed like Bayesian Methods, moving average and ad hoc methods like Fuzzy Logic and Neural Network used for multi-sensor data fusion. The scope of multi-sensor data fusion has to be defined. There are researches like Rodger (2012) where an integrated vehicle maintenance system was created using fuzzy Multi-Sensor Data Fusion (MSDF) to reduce failure risks. Two architectures are discussed here – Central Fusion and Distributed Fusion.

The study presents Fuzzy Kalman Filter Approach (FKFA) for better data error elimination as conventional Kalman Filter performs poorly under high computational number. The FKFA is used to reduce time needed to perform complex matrix manipulations in the vehicle maintenance model. If one system fails then, other sensors should take over the system to give collective decision for the same. Thus, a dynamic approach to selection of sensors should be present. This research uses dynamic detection and fault detection methods to find anomalies in the data of the vehicle. Another specific application of MSDF is discussed by De Paola, et. al, (2016) to cater to incorrect data because of inaccuracy in sensors. A Bayesian system in three-tier architecture of a smart home is proposed in the research. In the design, a dynamic group of sensors is used for data fusion to deduce inferences in a context aware environment. The lowest tier forms the sensor and reads all the environmental changes. The intermediate performs the multi-sensor data fusion and tries to infer the external conditions. The highest tier is the most complex where the system dynamics are modified to extract maximum features. This tier also caters to various factors like the cost and computational capacity and checks if they cross the predefined threshold. Dynamic Bayesian method used in this research create an information slice for a specified time while capturing the dynamic nature of past and present states. The designed system is tested in a controlled environment of a smart home. Reduction in energy consumption with redundancy is described using the technique. Compared to the Bayesian approach and FKFA described above, Gite & Agrawal (2016) describe another context aware process that uses sub processes like context acquisition, context processing, and context usage. The context processing contains noise removal, data calibration, context interpretation and context prediction. The research proposes Dempster – Shafer Theory (DST) in context-aware systems and compares the Bayesian approach with DST. It is theorized that when full probabilistic information is not available, DST performs better than Bayesian approach. This research proposes a technique but no implementation is done to prove the results. The computational capacity needed for the approach is also discussed which is more than the Bayesian Method. Here, with the study of researches, it is found that there are different approaches like FKFA, Dynamic Bayesian and DST for data fusion. These techniques all have their advantages and disadvantages depending on the application. The scope of these data fusion models is constrained to a small application where the data fusion of few sensors is done. Because of the scope, dynamic configuration is possible in the design and still the fusion complexity remains relatively low. The design is not scalable for a bigger number of systems working in tandem.

Along with specific applications like smart homes, research has also been done on larger scenarios like sensor data fusion for a city. The research Wang et al. (2016) gives comprehensive survey of different techniques and how the techniques can be scaled up for larger scenarios. It divides the sensor data fusion into three main categories. Complementary technique means putting together information from different sensors to get a bigger picture. Redundancy category fuses data from similar sensors to increase the accuracy of data. Cooperative operations combine to create new information. It describes combining different sensor networks and combines their data to form a predictive contextually rich model. This research also describes the evaluation framework for various data fusion techniques and compares various solutions based on the evaluation framework. This research does not propose any solution based on the findings of the study. Also, it also does not compare the computational cost and delay introduced in the communication.

Another comparison study of different techniques is given by Castanedo (2013) where the architectural analysis of different techniques is done. Castanedo (2013) describes the classification of data fusion techniques. It discusses the centralized, decentralized, distributed and hierarchical architecture of data fusion. In centralized fusion, all the data is concentrated in one particular central processor and then the data fusion takes place. Time delays are huge in this type of architecture. A node has its own computational capacity and there is no single point of data fusion in decentralized architecture. The communication costs in this architecture are high compared to centralized architecture. In distributed architecture, there are fusion nodes which fuse the data from several sensors. Thus, preliminary data fusion can be done at these nodes and brings down both computational cost and time of communication. Hierarchical architecture can be defined as hybrid model where data fusion is performed at different levels.

There are different techniques which are discussed here that have their own advantages and disadvantages. However, the scalability of the designs is not mentioned in any of the solutions along with other parameters like fusion complexity and implementation. There is a gap in the research to give a comprehensive survey of how the existing solutions can be scaled up to bigger architectures. This study intends to propose a design for a large scale scenario. Scalability of existing design will be studied, and next sections will explore the data fusion techniques in detail to offer a solution for data fusion.

## III. DESIGN AND FINDINGS

The proposed system is defined in this section using the knowledge gained in the previous sections. The architecture of data fusion in this design solution will be similar to that proposed by De Paola, et. al, (2016). This three-tier architecture as shown in Figure 3 completely covers all the architectural challenges. It is a distributed architecture as discussed above and can change to hybrid architecture depending on the sensors. The sensor subset varies according to the requirement of the system. This will bring in a dynamic environment which is needed if the system has to be scaled up.

The data association technique that will be used is the evidence reasoning method described above. Using this technique, the disadvantages of Nearest Neighbor method will be eliminated. This technique does not cost too much overhead in computational capacity and also increases the efficiency by more than 10%. The State Estimation method will employ the Fuzzy Kalman Filter Approach (FKFA). The ad hoc methods like Kalman filter do not perform well while putting it under high computations. Thus, FKFA will be beneficial regarding state estimation to predict current state of the system. The FKFA will bring in a dynamic environment and also remove the error due to interferences. The decision fusion method in the design solution will be implemented using Dempster -

Shafer method. The Dempster – Shafer method will predict the decisions and create a context-aware system. Thus, the design solution will provide a dynamic approach depending on the environment and also context-aware decision making.

The design solution will be shown to scale up and scale down depending on the application. The dynamic and context-aware system along with scalability is to be compared with existing platforms. The implementation strategy for this hierarchical architecture will be tough. Firstly, a dummy dataset should be verified with the data association technique. When the data association technique is tested, the state estimation and decision fusion methods are to be implemented. The self–optimization tier is the most complex to be implemented. This can be implemented on a computer using MATLAB software. MATLAB is efficient to implement methods like FKFA and Dempster – Shafer theory.

A detailed block diagram of the architecture is shown in figure 2. This block diagram consists of all the layers and the proposed solutions used for each stage in the design solution. The solution for each layer has been carefully selected to remove any chances of errors in the data fusion system.

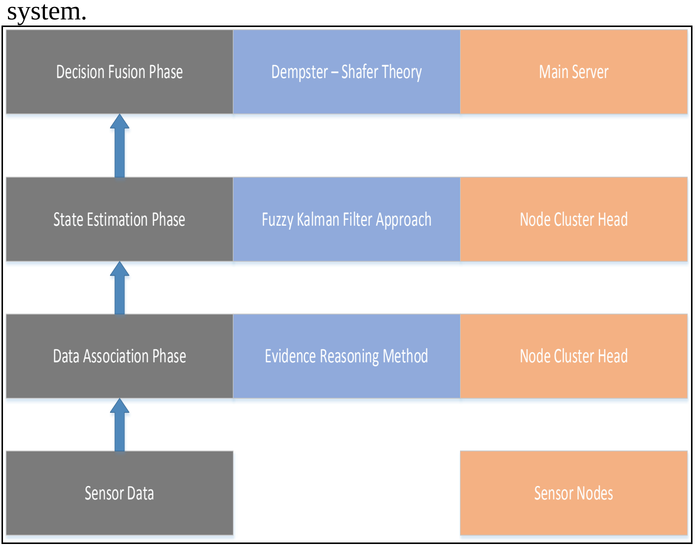

Fig 2: Final Block Diagram for the design solution

In this design solution, there is a main server similar to other models and a node cluster head to which all the sensor nodes are connected. In this design solution, the sensor nodes are given the least computation. As almost all sensor nodes are resource constrained, this removes any heavy computation load from the sensor node and it transmits the raw data to the node cluster. The node cluster is computationally advanced than the sensors, and it takes care of the decision fusion and state estimation for the sensor nodes. Thus, the Evidence Reasoning and Fuzzy Kalman Filter will be implemented in the node cluster head. The main server computes the decision fusion and eventually the data mining as well. Thus, the main inference is made by the main server. The main server is also responsible for the corrective measures that are to be taken. Thus, a hierarchical method for data fusion is proposed in this report. For implementation, the sensor nodes are implemented using small microcontroller units which measure different parameters like temperature, pollution level. The node cluster head can be implemented using an advanced embedded system which can communicate with the server and sensors, and perform data association and state estimation. The main server (cloud) can be simulated using simulation software.

There might be a possibility of errors in the implementation of this design solution, but it is hard to conclude without any implementation. The major area where errors might be generated is the central node where all the sensors are interfaced. The node cluster head forms the interface between two layers and also performs data association and state estimation. The resource constrained node cluster head would not perform well if many sensors are attached to it. Thus, for scalability, the cluster heads count will have to be increased which will increase the overall costs of the system.

### *Implementation and Performance Metrics*

The practical implementation of the proposed design can be done using emulators and simulation tools. The sensor nodes can be implemented using small microcontroller units with very less computation capacity. Sensors can be interfaced to these systems which might measure all types of parameters like temperature, pollution content, weather conditions, and traffic at a particular place. These sensors then send the compiled data to the node cluster head. The node cluster head contains the data association and state estimation phases. The master node is capable of changing the configuration of the sensors depending on the application and user data requirements. This feature of providing a dynamic configuration of the sensors will provide an added advantage to the system. The implementation of the cluster node head can be done using another high end embedded system which can control the dynamic configuration and send the data to the server. It also computes the data using Evidence Reasoning method and Fuzzy Kalman Filter Approach.

There are several performance metrics defined by Wang, et. al, (2016) used for scaling the design solution to a scale of a smart city. The parameters are listed below.

- Context Awareness – Combining data from various sources helps to understand the context properly.

- Dynamic Configuration – There are two types of dynamic configuration i.e. hardware configuration and software configuration. Hardware configuration is to provide multiple functionalities depending on the application. So, if a sensor is providing one type of data for some amount of time, then it should be hardware compatible to provide another data as well like temperature. Whereas, software configuration is to change the cluster configurations of the system and reuse similar functionalities to get different type of data.
- Type of Processing – Network processing and cloud processing of data are two types of data processing methods. A hybrid data processing method is used in this domain which processes the data in network as well as at the cloud level.
- Computation Cost – The total amount of processing that needs to be done on all levels of the architecture comprise of the computational costs. The computational costs at sensor level, network level, and cloud level should be considered. Ideally, the sensor should be provided with the least computation and cloud should be provided with the highest computation.
- Energy Consumption – Each stage of the architecture will use energy to compute the data. With massive amounts of nodes installed in various locations, the energy consumption needs to be kept at minimum.
- Scalability – The implementation strategy should be scaled from a small simulation to a big scale. Thus, for a smart city data fusion architecture scalability is of utmost importance.

### *Comparison of design solution with existing solution*

The design solutions discussed in various researches are compared with the design solution proposed in this report. Table 1 compares various implementations of data fusion to create a design metrics for implementation of data fusion methods. The design solutions discussed in the literature review do not cover all the performance metrics. Whereas, the design solution in this report covers all the features of the performance metrics and thus a better context aware and dynamic solution is proposed.

| **Parameter** | **(Rodger, 2012)** | **(De Paola, et. al, 2016)** | **Design Solution** |
|---|---|---|---|
| Context Awareness | Not available | Context aware system | Context awareness feature due to Dempster – Shafer Theory. |
| Dynamic Configuration | Fault tolerant and dynamic detection using software | Dynamically configures sensory infrastructure. | Dynamic Configuration possible using software configuration |
| Type of Processing | Hybrid processing | Cloud Processing | Hybrid processing using both network and cloud processing |
| Computational Costs | $O(m^2 + m \cdot n)$<br>Where, n = number of sensors<br>m = Number of possible values of various conditions in the system | Does not discuss computational costs. | Distributed computation shares the load of computation |
| Energy Consumption | Less energy consumption | High energy consumption due to computation of the data and transmitting the data wirelessly. | Less energy consumption as computation is done in cluster head nodes and main server. The sensors do not compute high amounts of data thus reducing the energy consumption |
| Scalability | Does not discuss any scalability feature | Solution is scalable depending upon subsystem discrete sensor needs. | Scalability is possible |

## IV. CONCLUSION

Data fusion is an old technique which is implemented in different domains of applications like image processing. The use of data fusion is not done in the Internet of Things domain and not popular. This research addresses the problem of lack of resources in the IoT domain for data fusion. In this research design, data fusion techniques have been studied to propose a unique hierarchical architecture for data fusion Internet of Things.

The design solution in this research proposes a new solution for data fusion. This unique solution combines the techniques of different implementations to provide a novel approach for data fusion. The design is robust with distributed computation on various layers of the solution which reduces the computational costs. Along with that, features like dynamic configurability and context awareness has been not be observed in other data fusion techniques together. The scalability feature is also included in the design which is described using an example of a supermarket and a smart city. A design metric is proposed to compare various data fusion algorithms before implementation. This design metric checks all the parameters which might have a significant effect on the system architecture. Thus, the design solution provides a novel approach for data fusion implementation in IoT domain and gives design metrics which can be used to compare the solutions with other implementations.

.

## REFERENCES

[1] Castanedo, F. (2013). A Review of Data Fusion Techniques. The Scientific World Journal, 2013, 1-19. http://dx.doi.org/10.1155/2013/704504

[2] De Paola, A., Ferraro, P., Gaglio, S., Lo Re, G., & Das, S. (2016). An Adaptive Bayesian System for Context – Aware Data Fusion in Smart Environments. IEEE Transactions On Mobile Computing, 1-1. http://dx.doi.org/10.1109/tmc.2016.25599158

[3] Gite, S. & Agrawal, H. (2016). On Context Awareness for Multisensor Data Fusion in IoT. Proceedings Of The Second International Conference On Computer And Communication Technologies, 85-93. http://dx.doi.org/10.1007/978-81-322-2526-3_10

[4] Koks, D. & Challa, S. (2003). An Introduction to Bayesian and Dempster-Shafer Data Fusion(pp. 1-48). Edinburgh: DSTO Systems Sciences Laboratory. Retrieved from http://ftp://www.ece.lsu.edu/pub/aravena/CoGS_Literature/Bayes&DempsterShaferDataFusion.pdf

[5] Rodger, J. (2012). Toward reducing failure risk in an integrated vehicle health maintenance system: A fuzzy multi-sensor data fusion Kalman filter approach for IVHMS. Expert Systems with Applications, 39(10), pp.9821-9836.

[6] Steinberg, A., Bowman, C., & White, F. (1998). Revisions to the JDL Data Fusion Model. In AIAA Missile Science Conference. § Monterey, CA: Naval Postgraduate School.

[7] Wang, M., Perera, C., Jayaraman, P., Zhang, M., Strazdins, P., Shyamsundar, R., & Ranjan, R. (2016). City Data Fusion: Sensor Data Fusion in Internet of Things. International Journal Of Distributed Systems And Technologies, 7(1), 15-36. http://dx.doi.org/10.4018/ijdst.2016010102

[8] Xiong, J., Zhang, Q., Peng, Z., Sun, G., & Cai, Y. (2016). Double sample data fusion method based on combination rules. IEEE Access, 99, 1-1. http://dx.doi.org/10.1109/ACCESS.2016.2604824.

## AUTHORS

**First Author** – Hemanth Kumar, B.E, M.Tech., M.S (Lancaster Univerisity) UK., Embedded Engineer, Ideal Educraft Technologies, Bangalore India, hemanthvk@gmail.com

**Second Author** –Pratik Pimparkar, B.E, M.Tech., M.S (Lancaster Univerisity) UK., Embedded Engineer, Ideal Educraft Technologies, Bangalore India, Pratik.pimparkar@gmail.com.