SMART TRANSPORTATION PLATFORM FOR BIG DATA ANALYTICS AND INTERCONNECTIVITY

Nandor Verba[1], Kuo-Ming Chao[2], Soizic Linford[3], Eleni Anoyrkati[4]
[1,2] Faculty of Engineering, Environment and Computing, Coventry University, United Kingdom
[3,4] Coventry University Enterprises Ltd, Coventry University, United Kingdom

Abstract: Innovations in Big Data Analytics and advances in Internet of Things have changed how we collect, interpret and analyses data from Smart City and Transportation environments. Advances in Fog Computing that suggests the extension of the Cloud to the Edge of the network using resources available to reduce latencies and improve reliability fit perfectly with concepts from autonomous and connected vehicles which have increasingly powerful processing capabilities. The use of such systems for congestion management, route planning, infrastructure management and other advances orchestration techniques requires the use of varying platform and technologies together. The existing proposals and platform are mostly designed to perform a fixed set of tasks, either purpose built or not generic enough to satisfy requirements from multiple use-cases. This paper analyses existing Fog Computing and Processing platforms characteristic data types and their brokers as well as existing Big Data cluster-computing frameworks like Apache Spark and Hadoop based on capabilities, data types and preferred use-cases. Supported by this analysis we propose a conceptual polyglot Fog Computing platform for Smart Transportation scenarios that can satisfy the needs of future smart city environments and can connect to multiple platforms and perform real-time, off line and hybrid processing of data as well. We provide a high-level framework for the integration of such platforms into collaborative clustering scenarios.

Keywords: Smart Transportation, Fog Computing, Big Data Analytics

1. Introduction

With rapid advances in autonomous vehicles and highly connected systems in the context of smart cities and transportation as well as the ubiquitous application of machine learning and data analytics in both local and cloud environments the field of future transportation and cities is rapidly expanding branching out into several research fields.

When considering Smart Transportation and smart, their interconnection in smart grid, opportunistic networks or clusters has become an increasingly appealing challenge. Especially with the introduction of concepts from Fog Computing with the local processing and decision making as an attempt to solve the scalability latency and some of the security issues inherent with cloud and centralized computing. This comes as there is an increased need for resource intensive processing and an increased concern over data-security and storage.

The recent Fog Computing framework proposals and IoT Middleware Reference architectures there is an increasingly clear consensus on the requirements of these gateways and of the use-cases in which they would be deployed. This makes possible the development of a system that can be can answer the required modularity, interconnectivity and lower level functionality abstraction to a degree where the development of management, control and analytics components for smart cities and vehicles becomes as trivial as mobile application development.

This paper proposes an IoT and Fog based Vehicle Centric Smart Transportation platform that look at the development of Smart Localities based on Wireless Sensor Networks (WSN) using Dynamic clusters based on fixed point gateways and mobile entities such as vehicles or other gateway enabled entities that cross these static regions and interact with them. Through this paper we will present a review of some of the existing middleware or gateway platforms, their advantages and drawbacks. Furthermore, we propose a dynamic application deployment and updating mechanism for these gateway for functionality extension as well as well as an experimental setup for the testing of the deployed systems.

2. State of the Art

When considering the background work that has been done for the development of these platforms there are several things that need to be considered. Concepts and Requirements need to be considered from the IoT domain that allow these gateways to be interconnected and for the connected devices or things to talk to each other in a protocol agnostic manner. When considering the added requirements and opportunities presented through Cyber-Physical systems a new set of interaction with the environment and users' needs to be considers together with an increased autonomy of the systems. These requirements are more dominant if more advanced approaches like Agent based Systems, Augmented Reality (AR) or autonomous vehicles are considered. These approaches require that the processing and other resources of the system be virtualizable and manageable to a level traditionally associated with Cloud computing. The paradigms and virtualization layers of new Fog computing approaches come into play in such systems, where some off the computational tasks can be offloaded to the edge, reducing costs and the strain on the network. These requirements and paradigms have led to several platforms being proposed that aim to solve portions of the problem. They range from Virtual Machine (VM), Container and shared environment solution, which offer varying levels of security and performance advantages.

---

[1] Corresponding author: verban@uni.coventry.ac.uk

2.1. Internet of Things

Internet of Things (IoT) is defined as the sensing analytics and visualization done anytime, anywhere and on anything (Gubbi, J. et.al.,2013) It can be considered the interconnected network of billions of devices as Gartner estimates at least 26 billion devices connected by 2020 (Rivera, J. et.al.,2013). This trend has reached the Through of Disillusionment on the Gartner Hype Cycle (Gartner Inc., 2015) and as of now a good view of the capabilities and the limitations of the technology are known and the wide-spread implementation and use of its components is being considered. When considering the application of its concepts for Smart Cities its main advantage is the seamless integration of a large set of heterogenous devices and systems (Zanella, A. et.al., 2014) When it comes to vehicle based systems, this can allow the Machine to Machine (M2M) communication of these through a variety of systems (P. *et al.2009*).

There are several views when it comes to the tasks of IoT devices. Some see the role of edge devices and systems simply that of data gathering or mining where (Almeida, R., et. al.,2017) looks at the networking aspect and (Chen, F.,et. al.,2015) at the semantics and components aspect of such an approach. Proposals as in (Gaur, A.,et.al. 2015) show the possibility of giving more agency to these edge devices by allowing decision making and connectivity tasks to be performed in a decentralized manor. The work done in (Datta, S.K., et.al. 2016) looks at communication with local networks while (Hasan, M., et. al. 2013) looks at random access or opportunistic networks being set up. The work in (Verba N. et. al., 2017) looks at more complex vehicular networks where the vehicles are the network creating a large peer to peer network while (Verba N. et. al., 2017) looks at their structure.

When considering the interconnectivity proposed by the Internet of Things, the field of Cyber-Physical Systems looks at how these systems should gather data, make decisions and interact with their environment. An overview of this can be seen in (Rawat, D.B., et. al. 2015) where they provide an overview of the impact of CPS on Smart vehicle systems and the impact of delays and latencies on these. The requirements of such systems can be seen in (Madakam, S., et. al.,2015) where a review of agent based systems is performed and a need for interoperability, flexibility and extensibility of such gateways which are in line with the IoT Requirements.

The processing, routing and mining of data for Traffic Monitoring, Management and Vehicle control is a crucial part of Smart Transportation systems as shown in (Chen, N., et. al., 2016) Typical IoT Systems deal with sensor and environmental data that are considered to have a medium frequency depending on the sampling and low in size when considering individual messages. The challenges of gathering and processing this data is mostly related to the quantity of messages and the semantics related to these sometimes, simple sensor messages

When considering all the research, it can be said that future IoT systems will require decentralized control and analytics within highly interconnected vehicles or devices that can interact with their surroundings through some sort of local or peer networks.

2.2. Fog Computing

Solving the problem of processing large data-sets on the edge while maintaining some of the capabilities and benefits of cloud systems was attempted by Cisco in (Bonomi, F., et. al.,2012) where they proposed the offloading of processing tasks to the edge of the network through Fog Computing. This approach aims to solve some of the latency, security and cost issues with using cloud systems, while improving on the scalability of a private clouds through the pre-processing of data locally.

The distributed control of such Smart Transportation systems has been proposed in (Banks, V.A., et.al., 2018) while (Hou, X., et. al. 2016) provides an overview of the addition of vehicles as system nodes forming the Vehicles as Infrastructure initiative. Proposals such as in (Hao, J., et. al. 2015) look at Urban planning based on big data management while (Hashem, I.A.T., et. al., 2016) looks at managing big data in the context of smart cities linked with industry, home and others. The researchers in (Burange, A.W. et. al.,2015) have proposed methods of increasing in security and privacy of data management through fog nodes as well. The proposal in (Suciu, G., et. al., 2013) attempts to use fog nodes as a means of increasing the reliability of existing systems while (Sanchez, L., et. al.,2014) aims to improve the scalability and impact of scalability.

Considering these proposals, Fog computing paradigms and approaches are being used to tackle some of the challenges posed by big data management and processing. Furthermore, consideration of scalability and security as well as platform lock-in are also important in similar systems.

2.3. Middleware and Gateway Platform Review

When considering gateways for smart city and transportation scenarios they need to satisfy the above-mentioned requirements and follow the existing directions if they are to be used as future platforms. An overview of some of the technologies and containers being used can be seen in (da Cruz, M.A., et. al. 2018) The overview of a few platforms implementing Fog Computing paradigms in the context of connected devices has be shown in Table 1. where they are analysed for their compliance based on the system requirements shown above.

The platform proposed in Urban IoT (Hao, J., et. al., 2015) provides a CoAP and generic HTML based IoT Environment for message routing with an analysis of acceptable delays in such systems, but fails to offer on gateway processing capability or any means of static or dynamic load distribution. The solution in IoT Smart City (Rathore, M.M., et. al., 2016) provides a polyglot, layered and hierarchical system view where the gateway has only a broker role. The data mining platform from (Chen, F., et. al.,2015) provides a higher-level data analysis for metadata extraction and semantics based on their layered app based architecture. The Smart Santander platform from (Sanchez, L., et. al.,2014) provides a large scale implemented network that is designed to facilitate future tests on IoT Systems. They provide a clustering and app based container that is also polyglot but don't offer abstraction or load balancing method within the apps. The proposal for Data Replication in (Verma, S., et. al., 2016) provides a way of deploying and managing Virtual Machines in a fog Server Tier. A similar proposal is suggested in (Bittencourt, L.F., et. al. 2015) that looks at high interconnectivity and migration support with device and message abstraction but without a cooperation mechanism.

Table 1
Overview of Fog Computing Platforms

Platform	**Inter-operability**	**Abstraction**	**Container**	**Cooperative Behavior**	**Load Distribution**
Urban IoT (Hao, J., et. al., 2015)	✓				
IoT Smart City (Rathore, M.M., et. al., 2016)	✓	✓			
Data Mining Platform (Chen, F., et. al.,2015)	✓	✓	✓		
Smart Santander (Sanchez, L., et. al.,2014)	✓		✓	✓	
Repl. Fog Env. (Verma, S., et. al., 2016)			✓	✓	✓
Cloudlet Migration (Bittencourt, L.F., et. al. 2015)	✓	✓	✓		✓

3. Vehicular Fog Computing Architecture

When formulating the components and choosing the protocols for the Vehicular Fog Architecture, the main directions, requirements and approaches that were used in the research reviewed in the previous sections need to be considered. One of the most important components that is most often noted in IoT systems is the capability to interconnect devices from different platforms and physical communication backgrounds. To do this our system needs to be able to allow components from different platform to interact using a common protocol or messaging service. The generated messages need to be translated or their protocol specific characteristics need to be abstracted away. These platforms need to be able to do local processing if the direction of Fog computing is to be followed and they need to be able to cooperate through service and resource sharing while allowing the use of these to be improved through either global or load balancing optimization.

Our approach is an extension of the work done in the PaaS approach from (Verba N., et.al.,2017) for such a system that can be seen in Fig. 1. where a high-level view of the system is shown. Here we can see that the proposed Platform design is deployed on each Node on the system, which can be a Home based, Car based, Cloud based, Personal and other types of nodes. When these are considered the main difference between nodes is they mobility. Some nodes are static nodes that either don't or rarely change their location and peer group. The mobile or dynamic nodes change their location constantly as they could be a personal phone, a car, or other vehicles in the system. These varying characteristics are used to create clusters with different interdependencies and connectivity characteristics.

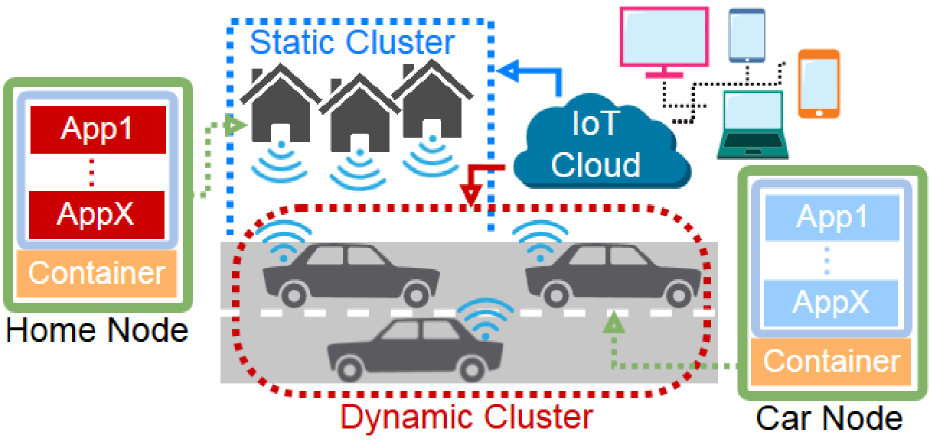

Fig. 1.
Overview of Smart Transportation Platform

The IoT Cloud connection layer is available to all peers in the system and is designed to allow for higher level data analysis and processing tasks to be run where local information and data is not enough or higher-level decisions need to be made. The cloud layer is designed to incorporate different protocols and link them with the existing analysis and decision systems.

3.1. Fog Platform

The Fog Platform Deployed on all nodes, both static and dynamic have at their core the platform described in Fig. 2. This element relies on the fundamental idea that all components should be accessed in a unified way and the system needs to separate events and control commands that are crucial to transmit from data streams that require high bandwidth but where low reliability and higher latencies are acceptable.

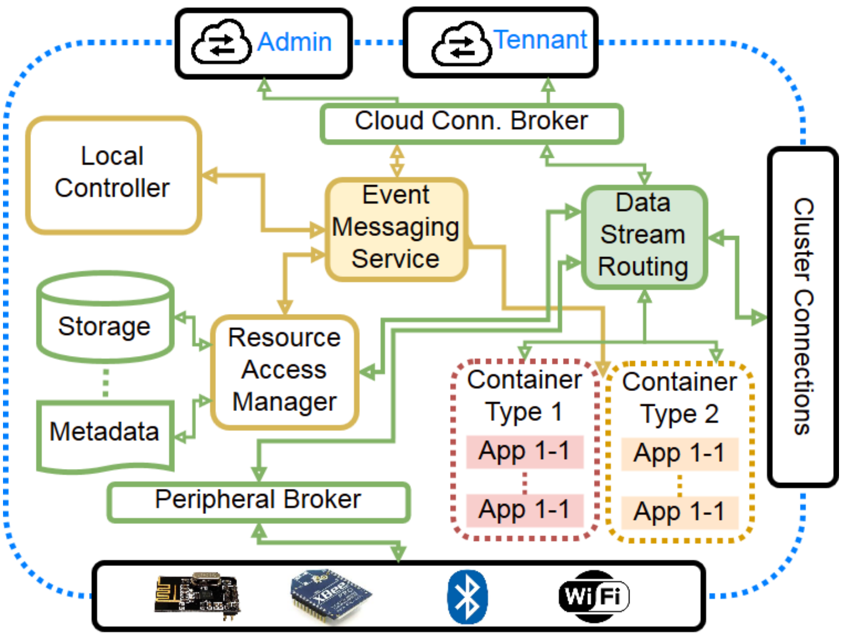

Fig. 2.
Overview of Smart Transportation Platform

The application containers on the system can vary based on use-case and requirements from, VM's to Docker images to Application deployed in shared environments. The main similarity between these is that they will have on message transmission system based on routing to send and receive information to peers. The same thing can be said for the peripheral, cloud and cluster brokers that take local events, messages and data then wrap these into the appropriate format for the sink they are sending it to. The Resource access manager controls the flow of information from apps and other peers to metadata information such as location, signal strengths etc. and varying storages. The local controller receives commands from on-site users and the administrative cloud which it uses to identify peers and control containers, the resource manager and the brokers.

3.2. Cloud Processing Layer

The Admin Cloud layer is tasked with performing centralized control commands on the deployed nodes and providing a user interface where the deployment schematics and system can be viewed. This layer is also tasked with authenticating nodes on the system and between each other. The management of users and region on these nodes is done locally. An overview of this and the tenant cloud layer can be seen in Fig .3.

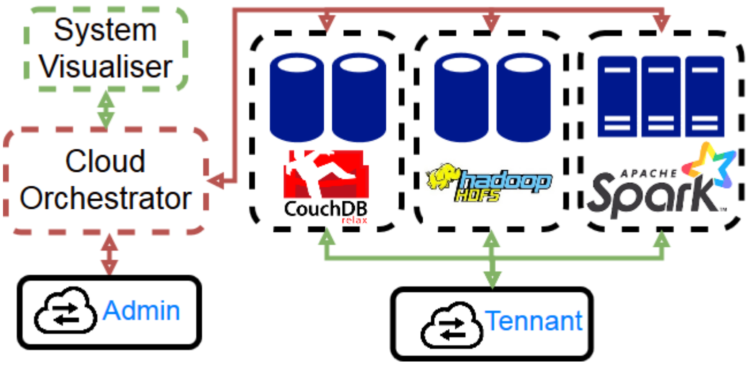

Fig. 3.
Cloud Orchestration and Processing Layer

The tenant cloud layer is designed to link the incoming data from application and nodes to the processing and storage units. The means and parameters of the storage are controlled by the cloud orchestrator which creates the databases and deployed the Hadoop or Spark tasks. This component is also responsible with configuring what happens to the resulting data.

The system visualizer or user interface allows managers and users of the system to have a graph based representation of the application deployed on the system and their interaction. This system aims to show at one time one node with all its applications and resources together with all the applications on other nodes it interacts with. This system aims to show all the apps that are deployed on other nodes but use information from the viewed node as well.

3.3. Clustering and Application Deployment

The creation of clusters and the temporary or permanent addition of peers is central to extending the functionality of IoT systems and realizing the requirements of Smart Cities and Transportation. The mechanism for clustering needs to be able to handle a highly heterogenous set of nodes and varying throughput of devices entering the system. For this to be achieved the authentication system needs to be able to gather information from the system and make decisions based on this. An overview of the connection phase and the subsequent service discovery, app request and system setup is shown in the sequence diagram from Fig.4.

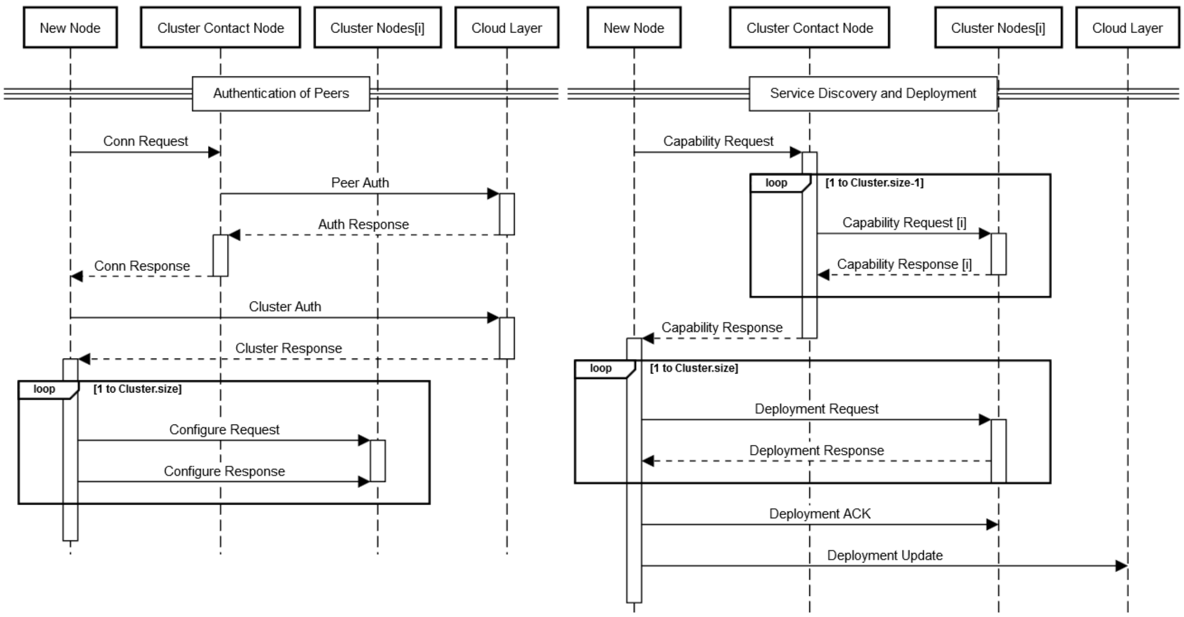

Fig. 4.
Authentication, Discovery and Deployment

The authentication process in based on simple two-way handshake between three entities where the cloud layer is the known and the safe one. The two entities verify each other in relation to the third one. This is done by the new node pushing a connection request to the first node in the cluster it discovered which is assigned a contact role. The contact node takes the information provided by the new node and verifies it through the cloud controller after which sends a response to the new node with the information it needs to authenticate the cluster and join. After the new node has then authenticated the cluster it joins in by sending the appropriate configuration requests to the new set of peers using the peer authentication parameters it received.

The Service Discovery and Deployment phase is designed for a new node to analyse the existing system and determine how it could extend the system functionality and how could the system extend the nodes functionality. The discovery is initiated by the new node by forwarding a request containing its own capabilities to the contact node which then queries the capabilities of its peers and responds to the new node by adding its own details as well. The new node than decides which apps it requires from the cluster and what endpoints and services it is interested in. It then sends the configuration request to each individual node, after which it configures itself. Finally, it sends a cluster wide acknowledgment to the nodes that it has successfully deployed everything after which it forwards and update of the system to the cloud layer.

4. Experimental Setup

The experimental setup is designed to test the functionality of the presented components and to allow us to evaluate how these behave in a real-life system. In the deployment we propose several typical household scenarios and evaluate how these can be deployed. Furthermore, the deployed gateways and devices are designed to be as heterogenous as possible, exploring several wireless communication technologies and protocols, as well as varying data-types, rates and applications.

4.1. Devices and Sensors

There were 7 types of devices used in the experimental setup in total. These range from the low capability Arduino and ATTiny based devices to the more powerful Nordic RF52 and to the Raspberry Pi Gateways carrying a varied set of sensors and actuators. An overview of these can be seen in Fig. 6.

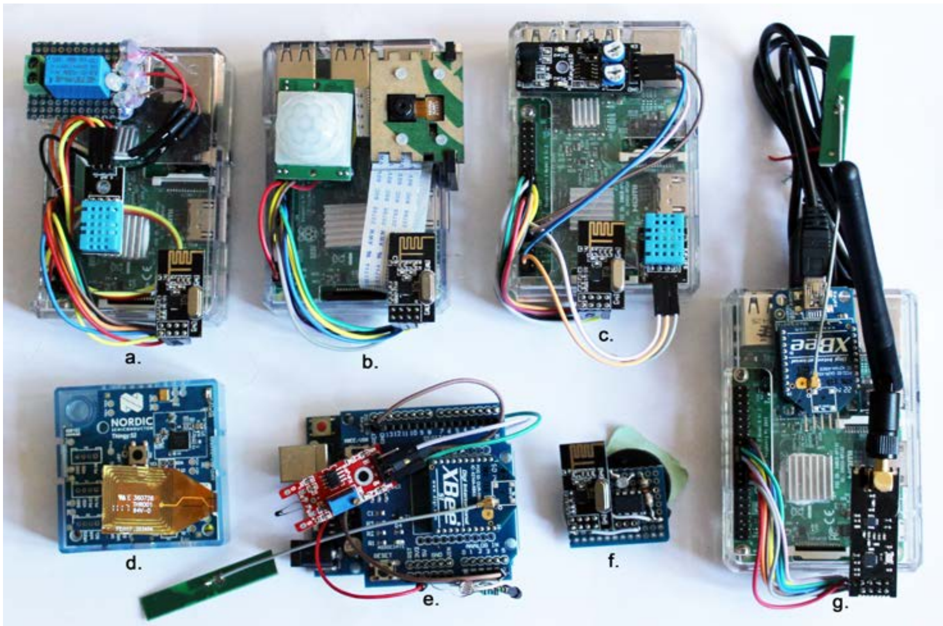

Fig. 6.
Experimental Devices

The Devices from Fig. 6 a,b,c and g are Raspberry pi 2 System on Chip Fog Nodes that have attached varying sensor that allow them to interact with their surroundings. Device a. has a relay, a temperature and humidity sensor and an RF24 wireless transmitter for communicating with peripheral devices. This is designed to control a boiler or a garage door controller while gathering environmental data and allowing access to devices. The Raspberry pi from b. is designed to be used with video surveillance, as it has attached a motion sensor and a Video Camera. It can be used for movement monitoring, tailored temperature control and for security reasons. The device c. has attached the temperature and humidity controller together with the RF24 module as well as a proximity sensor. It can be used as a door position monitor or car park monitor as well. The node in g. is designed to be a communication hub, as it can send and receive messages in all the testes technologies and has a higher range RF24 device.

The device d. is a standard RF52 Thingy node that has a varying set of environmental sensors and communication devices but is used to tests the Low-Power Bluetooth connections with the Raspberry Pi. It can also be used for actuation and visual alerts. The XBee enabled Arduino board from e. has a light and temperature sensor for environmental monitoring attached to it. The AtTiny device from f. is designed to be a simple monitoring device that can be powered using a 3.3 Volt battery and can have attached a light, temperature or proximity sensor.

5. Conclusions and Future Work

The deployment of IoT systems for Smart City scenarios requires a platform that can deal with the scalability and interconnectivity challenges that come from these systems. They need to be able to allow highly heterogeneous application and devices to communicate seamlessly. Furthermore, due to the large computational requirements these systems need to be able to efficiently use the resources available at the edge of the network.

We have reviewed a significant section of the existing work that has been done in the field concluding the requirements and directions. Based on this review we propose a platform that aims to solve these issues using Dynamic and Static clusters based on node characteristics. Furthermore, we propose a dynamic application deployment and updating mechanism for these gateways and an experimental setup for the testing of the deployed systems. The proposed experimental setup looks at providing nodes with highly heterogeneous capabilities, both from the communication, sensor capabilities and use case perspectives.

Our future work will look at large scale deployments of such systems with several dynamic devices that can be attached to the system and evaluation of bottlenecks on the system, how these can be circumvented and the limitations of the clustering approaches. We will also consider the deployment of multiple containers and scenarios using characteristic models from the literature and the optimization of these deployments.

Acknowledgements

Will have some stuff from prof. Chao Here.

References

Almeida, R., Oliveira, R., Sousa, D., Luis, M., Senna, C. and Sargento, S., 2017, December. A Multi-Technology Opportunistic Platform for Environmental Data Gathering on Smart Cities. In *Globecom Workshops (GC Wkshps), 2017 IEEE* (pp. 1-7). IEEE.

Banks, V.A., Stanton, N.A., Burnett, G. and Hermawati, S., 2018. Distributed Cognition on the road: Using EAST to explore future road transportation systems. *Applied Ergonomics*, *68*, pp.258-266.

Bittencourt, L.F., Lopes, M.M., Petri, I. and Rana, O.F., 2015, November. Towards virtual machine migration in fog computing. In *P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), 2015 10th International Conference on* (pp. 1-8). IEEE.

Bonomi, F., Milito, R., Zhu, J. and Addepalli, S., 2012, August. Fog computing and its role in the internet of things. In *Proceedings of the first edition of the MCC workshop on Mobile cloud computing* (pp. 13-16). ACM.

Burange, A.W. and Misalkar, H.D., 2015, March. Review of Internet of Things in development of smart cities with data management & privacy. In *Computer Engineering and Applications (ICACEA), 2015 International Conference on Advances in* (pp. 189-195). IEEE.

Chen, B. and Cheng, H.H., 2010. A review of the applications of agent technology in traffic and transportation systems. *IEEE Transactions on Intelligent Transportation Systems*, *11*(2), pp.485-497.

Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A.V. and Rong, X., 2015. Data mining for the internet of things: literature review and challenges. *International Journal of Distributed Sensor Networks*, *11*(8), p.431047.

Chen, N., Chen, Y., You, Y., Ling, H., Liang, P. and Zimmermann, R., 2016, April. Dynamic urban surveillance video stream processing using fog computing. In *Multimedia Big Data (BigMM), 2016 IEEE Second International Conference on* (pp. 105-112). IEEE.

da Cruz, M.A., Rodrigues, J.J.P., Al-Muhtadi, J., Korotaev, V.V. and de Albuquerque, V.H.C., 2018. A Reference Model for Internet of Things Middleware. *IEEE Internet of Things Journal*, *5*(2), pp.871-883.

Datta, S.K., Da Costa, R.P.F., Härri, J. and Bonnet, C., 2016, June. Integrating connected vehicles in Internet of Things ecosystems: Challenges and solutions. In *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2016 IEEE 17th International Symposium on A* (pp. 1-6). IEEE.

Gaur, A., Scotney, B., Parr, G. and McClean, S., 2015. Smart city architecture and its applications based on IoT. *Procedia computer science*, *52*, pp.1089-1094.

Gubbi, J., Buyya, R., Marusic, S. and Palaniswami, M., 2013. Internet of Things (IoT): A vision, architectural elements, and future directions. *Future generation computer systems*, *29*(7), pp.1645-1660.

Hao, J., Zhu, J. and Zhong, R., 2015. The rise of big data on urban studies and planning practices in China: Review and open research issues. *Journal of Urban Management*, *4*(2), pp.92-124.

Hasan, M., Hossain, E. and Niyato, D., 2013. Random access for machine-to-machine communication in LTE-advanced networks: issues and approaches. *IEEE communications Magazine*, *51*(6), pp.86-93.

Hashem, I.A.T., Chang, V., Anuar, N.B., Adewole, K., Yaqoob, I., Gani, A., Ahmed, E. and Chiroma, H., 2016. The role of big data in smart city. *International Journal of Information Management*, *36*(5), pp.748-758.

Hou, X., Li, Y., Chen, M., Wu, D., Jin, D. and Chen, S., 2016. Vehicular fog computing: A viewpoint of vehicles as the infrastructures. *IEEE Transactions on Vehicular Technology*, *65*(6), pp.3860-3873.

Gartner Inc., Gartner's hype cycle, 2015. http://www.gartner.com/technology/research/hype-cycles/.
Madakam, S., Ramaswamy, R. and Tripathi, S., 2015. Internet of Things (IoT): A literature review. *Journal of Computer and Communications*, *3*(05), p.164.

Papadimitratos, P., De La Fortelle, A., Evenssen, K., Brignolo, R. and Cosenza, S., 2009. Vehicular communication systems: Enabling technologies, applications, and future outlook on intelligent transportation. *IEEE communications magazine*, *47*(11).

Rathore, M.M., Ahmad, A., Paul, A. and Rho, S., 2016. Urban planning and building smart cities based on the internet of things using big data analytics. *Computer Networks*, *101*, pp.63-80.

Rawat, D.B., Bajracharya, C. and Yan, G., 2015, April. Towards intelligent transportation cyber-physical systems: Real-time computing and communications perspectives. In *SoutheastCon 2015* (pp. 1-6). IEEE.

Rivera, Janessa, and Rob van der Meulen. "Gartner says the internet of things installed base will grow to 26 billion units by 2020." *Stamford, conn., December* 12 (2013).

Sanchez, L., Muñoz, L., Galache, J.A., Sotres, P., Santana, J.R., Gutierrez, V., Ramdhany, R., Gluhak, A., Krco, S., Theodoridis, E. and Pfisterer, D., 2014. SmartSantander: IoT experimentation over a smart city testbed. *Computer Networks*, *61*, pp.217-238.

Suciu, G., Vulpe, A., Halunga, S., Fratu, O., Todoran, G. and Suciu, V., 2013, May. Smart cities built on resilient cloud computing and secure internet of things. In *Control Systems and Computer Science (CSCS), 2013 19th International Conference on* (pp. 513-518). IEEE.

Verba, N., Chao, K.M., James, A., Goldsmith, D., Fei, X. and Stan, S.D., 2017. Platform as a service gateway for the Fog of Things. *Advanced Engineering Informatics*, *33*, pp.243-257.

Verba, N., Chao, K.M., James, A., Lewandowski, J., Fei, X. and Tsai, C.F., 2017, November. Graph Analysis of Fog Computing Systems for Industry 4.0. In *2017 IEEE 14th International Conference on e-Business Engineering (ICEBE)* (pp. 46-53). IEEE.

Verma, S., Yadav, A.K., Motwani, D., Raw, R.S. and Singh, H.K., 2016, March. An efficient data replication and load balancing technique for fog computing environment. In *Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on* (pp. 2888-2895). IEEE.

Zanella, A., Bui, N., Castellani, A., Vangelista, L. and Zorzi, M., 2014. Internet of things for smart cities. *IEEE Internet of Things journal*, *1*(1), pp.22-32.