e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 7, July 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

6381 907 438 6381 907 438 ijmrset@gmail.com www.ijmrset.com

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707207 |

Resource Allocation Techniques in Edge/Fog Computing

Nirupama Kori, Raghu Prasad K

P.G Student, Department of Master of Computer Application, RNS Institute of Technology, Chansandra, Bangalore, India

Assistant Professor, Department of Master of Computer Application, RNS Institute of Technology, Chansandra, Bangalore, India

ABSTRACT: The advent of IOT has resulted in explosion of connected devices. These devices generally have limited computing capabilities and to overcome this there is a new distributed computing architecture in the form of edge computing. In this paradigm some part of the computation is performed at the edge of the network and the rest either in the end point nodes or the cloud depending on the task that needs to be completed. This method opens new challenges specially in the form of resource allocations in the edge computing framework. In this work we shall be looking into the major resource allocation paradigms present and new approaches to this problem Keywords - Internet of Things (IOT), Edge Computing, Cloud Computing, Fog Computing, DODAG, Distributed Computing, Resource Allocation Algorithms, Quality of Service.

I. INTRODUCTION

Cloud computing is fast gaining popularity as the preferred method of computation given the limited computing capabilities of end point node in IOT frameworks. However new use cases that require low latency and have strict Quality of Server Requirements (QOS) has resulted in a new computation paradigm, that of edge computing. In this paradigm majority of the computing resources reside in the edge of the network. Implementation of edge computing framework relies heavily on the method of virtualization like cloud computing and virtual machines are a in large aspects a key to success of edge computation. Resource allocation can be done as individual computing resources, or the concept of containers may also be used in this aspect. Containers are a virtual package consisting of computing and memory resources needed to complete execution of a process or task. A large degree of performance isolation is achieved using these. As per Cheol-Ho Hong et al [2], the main methodologies used for Resource allocation are as follows

1. Data Flow Architectures: The classification is made
based on the direction of data flow. Major models include
a) Aggregation
b) Sharing
c) Offloading
2. Control Architectures: The classification is based on mode of control. Types include
a) Centralized
b) Distributed
3. Tenancy Architectures: This classification is based on resource sharing and virtualizations
a) Discovery based: These class of algorithms try to find out the complete set of computing resources available in the edge framework. This is done based on different type of protocols including handshaking and messaging protocols. Security concerns of new devices joining the framework is a concern.
b) Computing performance benchmark: These algorithms compute performance benchmarks primarily with respect to power requirements, CPU and memory performance of edge processors
c) Load balancing: The tasks distributed should be sufficiently load balanced to gain maximum system efficiency.
d) Placement based: Identification of the appropriate resource for execution of a task.
Such allocations can be dynamic as well as static based allocation

The class of algorithms used for resource allocations are as follows

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707207 |

II. LITERATURE SURVEY & RELATED WORK

Edge computing is also referred to as Fog Computing. The computing module is a layer between the Edge Devices and the Cloud. The term Fog Computing was first Introduced by Cisco and the paradigm was introduced to handle computation offload from IOT devices. The Fog layer consists of a layer of switches ,controllers, computing, and storage nodes. As a concept it can be through of as an extension to cloud computing where the computation/ data storage wherever possible is carried out closer to the End Point IOT devices. This helps reduction of network latency and improves the network response timings. It might be mentioned in this aspect that Fog Computing is generally used as an addendum to Cloud Computing and not as a complete Replacement of Cloud Computing. The different layers in this computing model is as shown below.

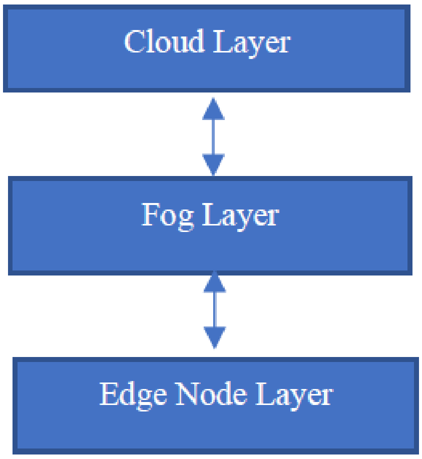


Figure 1: Layers in Edge/Fog Computing

The Fog computing architecture can be represented as in Figure 2. At one end we have a cloud unit and the edge node at the other end. In between the Cloud and Fog layer there is a cloud fog middle layer controller that arbitrates between the tasks that need be executed between the fog and the cloud layers [14]. The delay sensitive tasks are processed at the fog layer. The fog orchestration control node is responsible for allocation of resources to each of the fog nodes and performs the operation of a resource allocation manager for the nodes in a fog subsystem. The fog orchestration manager is generally one of the fog nodes. The fog nodes/cells are the computing nodes in the fog subsystem and are responsible for performing the task allocated to them.

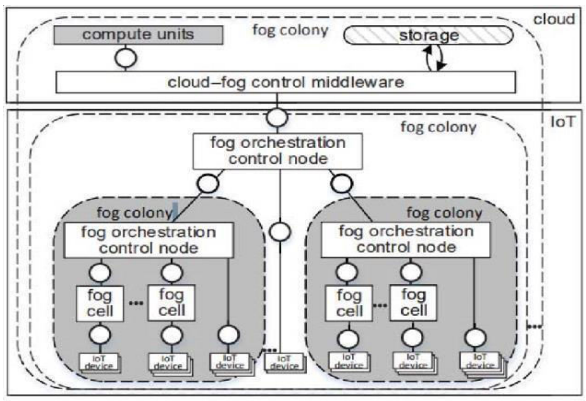


Figure 2: Fog Computing Architecture [15]

The fog nodes/cells receive a task request, performs the computation, and submits the results to another task node or back to the orchestration node. The computing power of a Fog node is measured by CPU MIPS, RAM, and the available network bandwidth. The storage capacity of the node may also be considered depending on the nature of the

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707207 |

task being performed. The complete Fog subsystem is dynamic nature and is to a large extent based on the paradigm of utility computing. New Fog computing nodes may be introduced and removed, and the task scheduling algorithm should be able to handle such dynamic changes. It might be noted in this aspect that the architecture described here is generic in nature and there might be implementations that do not have all the components mentioned here For example, we might have an implementation that is limited between the Fog and the Edge node layer and does not need a Cloud layer.

A fog node/cell details are presented in the following Figure3 [14] [15]

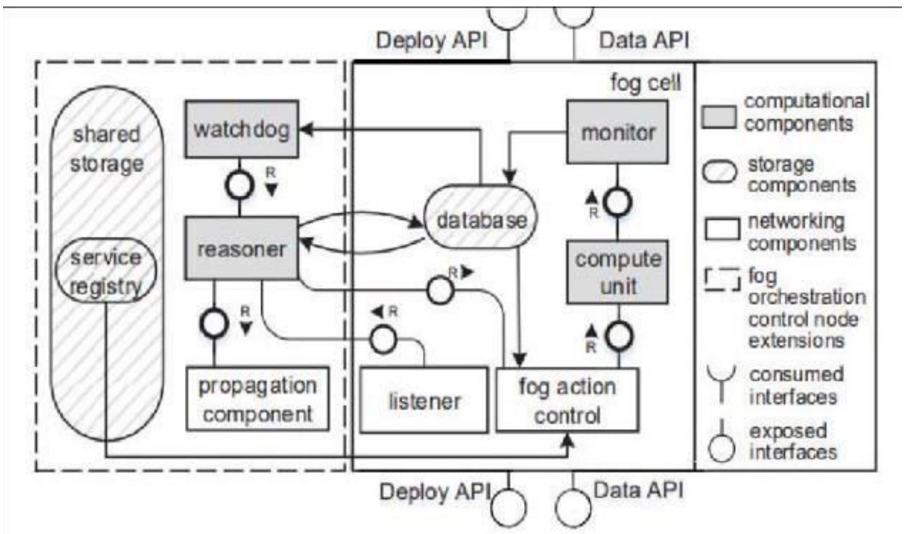


Figure 3: Fog Cell/Node Components [15]

For the remainder of this section we shall look at some of the prior arts in the field of resource allocation techniques in Fog/Edge Computing. Nguyen Quang-Hung et al [14] have explored a first fit algorithm for resource allocation. The computing nodes are represented in form of a tree and the first fit is allotted. The simulation is done using iFog Sim. Najamul et al [16] explores a Deep Reinforcement Learning method for determination of computation offloading to Fog devices/MEC servers. In DRL based algorithms generally Computing capacity (Processor MIPS and RAM/Caching) and network bandwidth are optimized. Agents learn about the network and ultimately determine an optimal offloading technique. The target was to develop a SDN controller to develop an adaptive policy for resource allocation that take into account the offload tasks for multiple users considering the mobility information. The target was to develop a joint optimization of battery consumption, network bandwidth utilization and computation resource optimization. As is multitude of other related algorithms a reward/penalty method is also utilized. Shuiguang Deng et al [17] propose a dynamic service resource allocation for distributed edges algorithm (DeraDE) that takes into account the time to a) transmit the data to the edge servers, b) allocate the correct optimal edge server, c? Execute the task and d) return the result back to the originating IOT device. Some of the other algorithms that are also effective for resource allocation comprise Locality frequency algorithm (LF), Low interreference recency set algorithm (LIRS) and Long shortterm memory algorithm (LSTM) Beomhan Baek et al [20] explore auction-based resource allocation algorithms including price based, uniform and fairness-based bidding algorithms In the following sections we propose an integrated algorithm that includes dynamic reservation of resources in the edge/fog computing nodes. The resource allocation is done in the form of a DODAG. Task Prediction modelling and Support for QOS is one of the major aspects of this algorithm. The proposed algorithm differs from others in its approach as it is based on the concept of pre-booking of resources in future point of time based on estimated current task completion time.

III. ALGORITHM

We denote the task profile in the form of an acyclic Graph G(V, E), where the set of vertices V denote the subtasks to be completed before completion of the task and the Edges denote the weight of the task involved in this process. The last node in this denotes the End Point node and can be assumed to have a constant weight.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707207 |

Task Call Flow:
a) All nodes in the Edge network are capable of conversing with each other and time taken to move data from one node to another is constant. Given the fact that the edge nodes are collocated this can be assumed without loss of generality
b) The target is to generate a DODAG based on the available information at that point. The DODAG (Destination Oriented Directed Acyclic Graph) contains the nodes needed to complete this task. It contains a sequence of nodes on which the task would be executed.
c) A protocol is already defined to circulate the capabilities of the nodes and an estimated load of the system is provided to all the nodes in the network. The protocol information element consists of
i. the history of the load in the node and its variations over the last 24 hours
ii. the capabilities of the computing node
iii. Current load of the node
d) The DODAG is computed at the task originating node. This node can be a constant and all tasks submitted to the Edge computing system can have a common ingress from which they are routed forward.
e) The DODAG can be recomputed at the next node based on the information available at that point of time and the best node is selected after computation of the current task is over.
Alternatively, we can also have static DODAG
f) The end of the DODAG is the End point that had requested for the task completion.
g) It is assumed that the subtasks comprising the main task are sequential in nature. The challenge is to select the next most suitable machine to accomplish the next subtask
h) The given task needs to be computed across data set placed across multiple machines but due to concerns related to security data duplication is likely to be limited and very controlled
i) It is implicitly assumed that the machine that has the data is used for computation
j) The creation of the DODAG helps eliminating any possible loops that may be seen during routing of the tasks as the path is generated beforehand and validated during execution. The egress node is determined beforehand thereby eliminating any possible loops

Protocol Call Flow:
a) The edge computing nodes are connected over a local LAN
b) The edge nodes follow a publisher/subscriber model to publish computing/capability information among each other. A publisher/subscriber model is a typical IOT based information transfer protocol and is supported by most IOT protocols
c) Any change in resource allocation/deallocation due to change in DODAG route is propagated as well.

Support for QOS:
a) The model takes care of QOS requirements of the desired task.
b) During establishment of the DODAG, a bandwidth is reserved in the computing nodes based on Estimated time of completion of the previous tasklets in the DODAG
c) An Estimated Task arrival time and Task completion time is calculated beforehand while
creating the DoDAG
d) A change in computation completion time and revised timings are sent down the DODAG path to adjust any timings for bandwidth reservations. This helps to adjust with the dynamic changes in the subsystem
e) The method of calculation of estimated arrival time and completion time based on existing load and new load arrivals. The prediction model can be based out of machine learning and is a target for our future scope of works

Task Prediction Model
The task offload and Edge/Fog node selection is based on a model that is based on the Task load prediction of the network. When selecting the appropriate node for task execution both the current load as well as the Futuristic load is considered, and network and computation bandwidth is allocated based on that. This is calculated and validated at each node. If the latest calculated route path is better than the original calculated path by a specific threshold then only a rerouting of tasks is considered. In our subsequent research work we plan to study the effects of
a) Varying weights of Futuristic load on the prediction model and its efficacy
b) Different distributions of futuristic load
c) The choice of threshold that determines whether it is better to continue with the existing path or the task needs to be re-routed

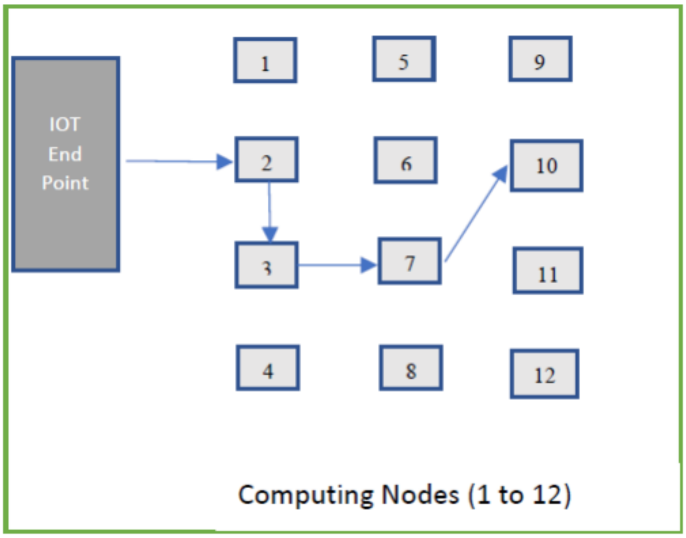
As part of this model, resources at the designated nodes in the current DODAG are booked so that when the task is finally executed at the designated node, the wait time is minimized. Any re-planning of the DODAG will result in

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707207 |

releasing the acquired resources and booking of a fresh resource at future time slots based on the current and futuristic load in the system. It might be noted in this aspect that any re-planning of the Task Path (DODAG) would have an overhead as that would mean allocation deallocation of bandwidth/resources at multiple nodes and information propagation of the same across the computation framework. This makes the choice of threshold to determine a re-planning or not very critical.



In the above sample model, the IOT Endpoint submits thetask at node2. Node 2 can be a fixed ingress point into the Computing Mesh nodes or it can be a chosen representative based on a election among the peer nodes. The task is then routed along the nodes {2 →3 →7 →10}

before it is submitted back to the End Point node. The DoDAG is calculated at the entry point node 2 and includes the graph {2 →3 →7 →10}. At each node the DODAG a sanity check is done with respect to the availability of the node and the computation proceeds. This helps in reacting to changes in the network. The decision of the selected route is taken based on the Task Prediction Model discussed in the previous section

IV. CONCLUSION

In this paper we had a look at some of the major classes of resource allocation algorithms in edge computing and proposed a Discovery based Graph scheduling algorithm. The proposed algorithm uses a DODAG based approach to represent the task flow and chooses the best node at every step. Such an algorithm can be used when the given task needs to be computed across data set placed across multiple machines but due to concerns related to security data duplication is likely to be limited and very controlled. Another assumption is that the machines that have the data are the only ones that can perform the computation operation. This algorithm is capable of handling QOS demands of the subsystem and is based on future time resource allocations in subsystems. The efficiency of the prediction of time of task arrival and completion is core to the efficiency of the subsystem.

V. FURTURE WORK

In the current approach discussed, to be able to meet the QOS requirements a model of advanced booking of resources at the different task nodes in considered. Pre booked resources means that in the whole task flow, waiting time for a task can be minimized. We plan to study the effects of wait time that a task has based on thevarious parameters of this "Future Resource" booking algorithm. One of the assumptions of the current model is sequential execution of tasks. As a part of future work, we would explore the possibilities of parallel execution of such tasks.

REFERENCES

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707207 |

[1]A Novel Bio-Inspired Hybrid Algorithm (NBIHA) for Efficient Resource Management in Fog Computing by Hina Rafique , Munam Ali Shah , Saif Ul Islam, Tahir Maqsood, Suleman Khan, and Carsten Maple, Special Section On Mobile Edge Computing And Mobile Cloud Computing: Addressing Heterogeneity And Energy Issues Of Compute And Network Resources, Volume 7, 2019

[2] Resource Management in Fog/Edge Computing: A Survey on Architectures, Infrastructure, and Algorithms by CHEOL-HO HONG and BLESSON VARGHESE, ACM Computing Surveys, Vol. 1, No. 1, Article 1, Publication date: September 2019.

[3] Architectural Framework for Resource management in Fog Computing by B J. Hubert Shanthan, Dr. L. Arockiam, International Journal of Engineering Research & Technology Authorized licensed use limited to: East Carolina University. Downloaded on June 29,2021 at 15:05:43 UTC from IEEE Xplore. Restrictions apply. (IJERT) ISSN: 2278-0181 Published by, www.ijert.org ICONNECT - 2017 Conference Proceedings

[4] Dynamic Resource Allocation for Load Balancing in Fog Environment, Xiaolong Xu, Shucun Fu, Qing Cai, Wei Tian, Wenjie Liu , Wanchun Dou , Xingming Sun , and Alex X. Liu, Hindawi Wireless Communications and Mobile Computing Volume 2018, Article ID 6421607

[5] Efficient Resource Allocation for Time-Sensitive IoT Applications in Cloud and Fog Environments by Lina Husain Kazem, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-3, September 2019.

[6] Simulating Fog and Edge Computing Scenarios: An Overview and Research Challenges Sergej Svorobej, Patricia Takako Endo, Malika Bendechache, Christos Filelis-Papadopoulos, Konstantinos M. Giannoutakis, George A. Gravvanis , Dimitrios Tzovaras, James Byrne and Theo Lynn, Future Internet 2019, 11, 55

[7] An overview of service placement problem in Fog and Edge Computing, Farah Ait Salaht, Frédéric Desprez, Adrien Lebre, [Research Report] RR-9295, Univ Lyon, EnsL, UCBL, CNRS, Inria, LIP, LYON, France. 2019, pp.1-43.

[8] TACRM: trust access control and resource management mechanism in fog computing by Wided Ben Daoud, Mohammad S. Obaidat, Amel Meddeb-Makhlouf, Faouzi Zarai and Kuei-Fang Hsiao [9] Edge Cloud Infrastructure for the future network by Yuang Zhang 21-5-2019

[10] Towards Distributed Data Management in Fog Computing by Vasileios Moysiadis, Panagiotis Sarigiannidis , 1 and Ioannis Moscholios by Hindawi Wireless Communications and Mobile Computing Volume 2018, Article ID 7597686, 14 pages https://doi.org/10.1155/2018/7597686

[11] Methods And Techniques To Deal With Big Data Analytics And Challenges In Cloud Computing Environment by Dr. S .Mohan Kumar Naga Raju Hari Manikyam, International Journal of Civil Engineering & Technology, 2017

INNO SPACE
SJIF Scientific Journal Impact Factor

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |