Vision for Research

The advent of interconnected smart and intelligent systems has enabled efficient delivery of different kinds of services in areas such as transportation, smart homes, smart grid, medical and healthcare, etc. Such a large conglomeration of connected devices/entities form the Internet of Things (IoT). The large scale integration of heterogenous entities in an IoT brings with it the inherent complexity of system level design and analysis. In addition, these IoT systems are comprised of a large number of real-time embedded systems, which require schedulability analysis and design techniques specific to real-time systems. My primary research interests address these challenges and fall under the broad theme of "**Scalable Design and Performance Analysis of IoT and Real-Time Embedded Systems**", especially delving into topics like *edge computing based services for intelligent transportation* and/or *safety critical real-time embedded computing systems for automotives*.

One of the major challenges in performance analysis of applications running in IoT and Real-Time Embedded (RTE) systems is the predictability of performance objectives. Predictability in RTE systems is influenced by the variability in available computation, communication and memory resources for processing an application as a result of contention on the resources from competing tasks. However, the RTE system is typically confined in space and dimension and its performance analysis is not influenced by network communication technology. On the other hand, the performance analysis of an IoT system needs to consider a larger system view involving network communication technology and possibly user input devices in addition to the RTE sub-systems. In this context of both systems mentioned above, I envision to work towards proposing novel robust and efficient design and analysis techniques for *timing predictability*, *energy/thermal efficiency*, *system resilience* and *optimal resource usage*.

1 Prior Work

1.1 Performance Analysis, Scheduling and Design of RTE Systems

During my doctoral study, I developed several formal analysis techniques to perform worst/average-case analysis in various problem contexts pertaining to executing hard/soft real-time applications on multiprocessor system-on-chips (MPSoCs). The three techniques developed had its theoretical roots in Real-Time Calculus (RTC) (a well known tool for compositional analysis of distributed systems) and partly also in resource reservation-based scheduling [1, 3, 4]. Further, I also developed analytical techniques based on Stochastic Network Calculus to derive quality-of-service (QoS) for execution of multimedia streams on MPSoCs quantified as probabilistic bounds on performance [7, 8]. In addition, I worked on techniques to perform fast hybrid simulation [5, 6] for performance analysis of multimedia MPSoC platforms. The problems that I solved during my PhD gave me the necessary skill set to rigorously analyze non-functional aspects of executing applications on embedded platforms.

As a PostDoctoral Researcher, I worked on several problems delving into proposing new design techniques for better timing predictability of executing concurrent applications on MPSoC platforms. Some of these works were based on the concepts of composability and hybrid design space exploration (DSE) of NoC-based multi-tile platforms using spatial isolation [9] and temporal isolation [10] of shared resources. One line of work also addressed the issue of contention on buses by firstly developing a method to reconfigure the AMBA AHB bus at runtime [11] and then proposing an online algorithm to perform reconfiguration of bus scheduling policy between non-preemptive fixed priority (NPFP) scheduling to time division multiple access (TDMA) scheduling with low complexity [12]. These works on embedded system predictability gave me very good insights regarding the influence of hardware platform and system software overheads on the performance of execution of real-time applications.

As part of the EU ARTEMIS project, I worked on the development of DSE techniques for Application-Specific Instruction Set Processor (ASIP) based multiprocessor platforms. The challenge was that the performance of an ASIP-based design can be done only when the worst-case execution time of the application tasks mapped to ASIPs are known, which requires knowledge of ASIP microarchitectural details, but the microarchitectural details cannot be decided until the mapping of application tasks on ASIPs is finalized. This results in a cyclic problem of how to perform ASIP macro level DSE, which was addressed using a probabilistic DSE method leveraging a probabilistic WCET modeling of application tasks. This led to several publications in Real-Time Systems conferences [22, 23] and a journal [24] in systems. These works further expanded my knowledge of multiprocessor platforms by considering the very important microarchitectural details.

In the domain of real-time scheduling, one of my papers [13] at IEEE ISORC 2018 (won the **Best Paper Award**) (co-authored with a graduate student I mentored at University of Pennsylvania) proposed a method to determine the periods of tasks in chains of arbitrary length while satisfying end-to-end data freshness constraints with only few assumptions regarding the scheduling algorithm used. The novelty of the approach was the use of a scheduler agnostic method for determination of periods, which helps in the analysis of a system at an early stage when the platform scheduling details are unknown. However, this work has lot of potential in terms of developing algorithms to ensure the end-to-end data freshness requirements while executing periodic applications on safety critical hardware platforms.

1.2 Analysis and Design of Edge Computing-based Intelligent Connected systems

In my current role as a faculty at IIIT Hyderabad, I have worked on efficient analysis and system design solutions for complex edge-based Intelligent Connected Systems.

A very important scenario in this category is a connected vehicle and my team has explored novel algorithms to deliver data/services to connected vehicles via the edge node while the vehicle is on the move. This requires allocation of adequate amount of memory, computation and communication resources on the edge nodes. My first work in this area [15] formulated the data/service delivery problem in a vehicular edge computing system as an optimization problem, which minimizes the system wide total bandwidth cost of the edge nodes. As a follow up work, I proposed a social welfare based optimization framework [16] for data/service delivery considering both delivery time and total edge bandwidth cost.

My team of 2 MS students have worked on three very important problems on efficient data/service delivery heuristic algorithms for connected vehicles via edge nodes when they are on the move. The first work proposed a fast heuristic which delivered data/service and was able to service a large number of vehicles very close to the optimal approach (which was published in **IEEE Cloud 2018**) with very less departure from edge bandwidth cost optimality and significant orders of improvement in the time required to deliver data/service. This paper was published in the **IEEE Vehicular Technology Conference (VTC) 2021** [17]. The above work was further improved by proposing a novel approach to determine the overlap of vehicles within the coverage area of an edge node. The vehicle overlaps were not considered in the previous work and the assumption was that all the vehicles passing through an edge's coverage region does it during the same time, which is the worst-case assumption. This resulted in over allocation of edge resources to the vehicles resulting in servicing of lesser number of vehicles compared to the optimal approach. Our proposed method [18] including estimation of vehicle overlaps enabled the service of more vehicles and improved the earlier work [17]. Further, we recently proposed data/service delivery frameworks via edge nodes while considering channel contention between vehicles and the edge [27] and for dynamic variation in the route for the vehicles [28].

The use of edge hardware platforms to perform complex machine learning or deep learning tasks is on the rise. We have attempted to leverage this for the domain of Intelligent Transportation. One of the first works in this area looked at two-wheeler safety. There has been a lot of research in advancing the safety of four-wheelers due to rising interest in self driving cars, but similar advancements have eluded the two-wheeler space. As the number of two-wheelers is significantly high in many Asian countries, it is necessary to provide some safety functionalities that can be executed in the micro-controller within the two-wheeler. We are working on developing efficient deep learning algorithms to detect accidents and bike driving events. These algorithms have been deployed on an edge device, which will detect accidents or driving events in real time. Specifically, we developed the LSTM and Bi-LSTM models with attention mechanism, which gave the best accuracy of 92% over state-of-the-art models [25] for detecting bike accident in 2-3 seconds. We have also developed the time-series data based classification of driving events, which was presented at **DATE 2023** [19]. The idea behind classification of driving events is to derive driving quality index to quantify the driving quality of a rider.

In the past couple of years, my team has also been looking at developing a edge-cloud architecture with fail safe mechanism to predict traffic flow prediction in disorganized traffic, which is very prevalent in some of the Asian countries. We have developed temporal traffic flow prediction models that have been deployed on Raspberry PI and predicts short term traffic flow of a local intersection [26]. The cloud looks at more spatial data along with the temporal data in a location to give more accurate flow estimation. The predicted flow value can be used to perform traffic flow control, which is currently being studied.

1.3 Development/Analysis of IoT Systems with Interoperability Middleware Layer

After joining IIIT-H, I joined the team of Smart City Living Lab project and took on the leadership role as the faculty responsible for the development of software infrastructure to enable the usage of an IoT Middleware for inter- operability (based on a widely used oneM2M standard) of the different IoT applications/verticals. We are using an open source implementation of the oneM2M standard called OM2M. The initial challenges were to improve the performance of data storage and retrieval from the OM2M implementation. After achieving some improvement, we worked on the integration of the OM2M implementation with an external database Postgres. We have proposed a novel multilayer architecture and evaluated the performance of the architecture with varying request frequencies and number of requests on the real testbed in the IIIT-H Smart City Lab, which has been published in **IEEE WF-IoT 2022** [21]. This was further improved in terms of performance by using parallel instances of the interoperability layers along with its local database and published in **IEEE WF-IoT 2023** [29]. We have also integrated the OM2M-Postgres database setup with a data exchange framework called IUDX that will be used for Smart Cities in India. I have also worked on the development of IoT nodes that measure the TDS value of water in RO tanks and sump in the IIIT-H campus so as to keep track of the quality of water. As part of this work, we published a paper in the reputed conference **IEEE WF-IoT 2021** [20] which proposed a non invasive technique of measuring TDS values of water in RO tanks. The TDS sensor was also calibrated using machine learning.

2 Future Research Plan

2.1 Performance Analysis, Scheduling and Design of RTE Systems

Under this research thread, my future work will delve into problems in the space of scheduling, analysis and design of RTE systems considering the following:

1. Emerging hardware platforms such as 3-D stacked processor memory system, which is increasingly being looked at due to their high performance.
2. Many core platforms with heterogeneous processors.

The main problems in scheduling and analysis of RTE systems in the above platforms is to derive theoretical and practical insights into the satisfiability of requirements of non-functional objectives such as *data freshness*, *security over shared resources*, *energy/thermal efficiency* and *resilience*. The above objectives are of utmost importance in the context of RTE systems relevant to automotive platforms or other safety critical systems such as unmanned aerial vehicles.

Some of the interesting research problems that will be explored within this theme are described below.

1. Formalization of the theory to map and schedule periodic task chain based applications (or cause-effect chains) on the above mentioned platforms such that data freshness can be optimized. This is a novel problem as there is not much work that performs a data freshness optimized deployment of such applications on multi or many core platforms as mentioned above.
2. A formal approach to analyze and schedule cause-effect chains in a fault tolerant manner considering checkpointing and redundancy within the above mentioned platforms while satisfying end-to-end data freshness constraints. Fault tolerance has been studied extensively in the context of system schedulability, but there is very little which looks at fault tolerance strategy while satisfying data freshness.
3. Analysis of mapping and execution of cause-effect chains in the above mentioned platforms while satisfying an end-to-end data freshness constraint under energy or peak temperature minimization. The conventional energy and power budgeting mechanisms can be put to test here and we can explore its limitations while considering data freshness.
4. Analysis of mapping and execution of cause-effect chains on the above platforms and the effect on end-to-end data freshness with limited preemption schedulers. This is a novel problem as limited preemption scheduling and its effect has only been explored in the context of system schedulability.
5. Security-aware scheduling [14] is an important research direction, which has many unanswered problems. Within this direction, I would like to propose analysis and scheduling techniques that will optimize the end-to-end data freshness, while providing the required level of security to the applications.

Preliminary versions of the above ideas are currently being performed under the seed grant provided by IIIT Hyderabad. There is also work going on in the area of fault-tolerant control and security-aware scheduling for UAVs, which is supported by two grants from the Department of Science and Technology (DST), India.

2.2 Analysis and Design of Edge Computing-based Intelligent Connected systems

Data/service delivery in Vehicular Edge Computing (VEC) is an ongoing area of research as there are different kinds of services and data that are required to be delivered to the vehicles to enhance the driving experience. With increasing scale of VEC networks and several distinct system objectives to be optimized, there are many problem scenarios where resource allocation on the edge is an important problem. The open problems in this space that will be addressed in future are

1. **Security-aware data/service delivery to vehicles via edges**: The current resource allocation algorithm on edges for delivery of data/service to vehicles has not considered the possibility of attacks that can tamper with the data or corrupt the service. This can have serious consequences if some critical data or functionality is attacked. As per the Automotive Edge Computing consortium (AECC), there are multiple levels of security that can be provided based on the application scenario. In this context, I will be exploring online resource allocation algorithms that will provide adequate security to vehicle services while optimizing some system objectives.
2. **Scheduling Algorithms to deliver data/service to vehicles**: We are exploring the resource allocation problem as a multi-core scheduling problem considering partitioned and global scheduling strategies, while considering vehicle flow constraints and edge bandwidth constraints. The novelty of this approach is the transformation of the resource allocation problem in a VEC context to a well known multi-core scheduling problem.
3. **Learning based content caching algorithms in VEC**: The state-of-the-art in this work looks at reinforcement learning techniques to allocate a certain amount of cache units for each vehicle on an edge node, but does not consider the vehicle flow constraints and edge bandwidth constraints. The efficient way to cache content is by allocating variable bandwidth on the edge nodes. We will develop a reinforcement learning strategy to realize content caching considering variable edge bandwidth allocation.
4. **Reconfigurable hardware design to realize the data/service delivery**: The workload on the edge nodes is becoming complex and heterogeneous considering the different kinds of services that the vehicles will demand in future. These could be safety critical or non critical tasks. The variability of this workload also can keep changing over a period of time. Considering these factors, it is necessary to design a hardware platform with heterogeneous resources consisting of reconfigurable elements. This will also give the flexibility to decide at run time what workload should be executed on which resource. Hence, the work will involve the development of hardware and algorithms to enable the execution of varying service requirements of the vehicle.

In addition to data/service delivery for connected vehicles, one other area that I have started exploring is the use of deep learning (DL) algorithms for connected vehicle applications using the edge device. As DL based edge analytics is performed on resource constrained edge devices, I see the potential to conduct research on DL algorithms that use intelligent information generation and sharing mechanisms to optimize the energy efficiency and thermal requirements of the edge device. Currently, there is significant amount of research on developing efficient DL algorithms to execute on the resource constrained edge devices. One direction in this research is to trade off accuracy of the DL inference by pruning some feature data and thereby reducing resource usage and time. We are exploring this in a research project looking at real-time efficient traffic flow prediction on Indian city roads using edge nodes. If there are safety critical automotive tasks that need to be executed on the edge device (for e.g., offloaded tasks from vehicles) along with the DL inference task, a pertinent problem will be to analyze how much accuracy of the DL task needs to be traded off with the timeliness of the safety critical tasks, in the worst-case, given an arrival pattern of the safety critical tasks. **Currently, all the above works are partly being conducted with a generous grant from the Department of Science and Technology (DST), India under the National Mission on Interdisciplinary CPS**.

Vehicle Platoon formation is another scenario where there is a lot of potential of using edge computing and deep learning algorithms. Most of the platoon formation decisions are performed either offline or with very limited local information. A very important challenge in forming efficient platoons is the dynamic nature of traffic making it very difficult to use the offline decisions. Therefore, data collected by edge nodes can be used to make intelligent decisions regarding the optimal platoon formations considering traffic flows and vehicle routes. The grand idea here is to develop algorithms and system to facilitate real-time efficient platoon formations for logistics companies.

References

[1] **D. Gangadharan**, L. T. X. Phan, S. Chakraborty, R. Zimmermann, I. Lee, "Video quality driven buffer sizing via frame drops", *RTCSA*, 2011, pp. 319-328.

[2] **D. Gangadharan**, H. Ma, S. Chakraborty, R. Zimmermann, "Video quality-driven buffer dimensioning in MPSoC platforms via prioritized frame drops", *ICCD*, 2011, pp. 247-252.

[3] **D. Gangadharan**, S. Chakraborty, R. Zimmermann, "Quality-aware media scheduling on MPSoC platforms", *DATE*, 2013, pp. 976-981.

[4] **D. Gangadharan**, S. Chakraborty, J. Teich, "Quality-aware video decoding on thermally-constrained MPSoC platforms", *ASAP*, 2014, pp. 256-263.

[5] **D. Gangadharan**, S. Chakraborty, R. Zimmermann, "Fast model-based test case classification for performance analysis of multimedia MPSoC platforms", *CODES+ISSS*, 2009, pp. 413-422.

[6] **D. Gangadharan**, S. Chakraborty, R. Zimmermann, "Fast hybrid simulation for accurate decoded video quality assessment on MPSoC platforms with resource constraints", *ASP-DAC*, 2011, pp. 237-242.

[7] B. Raman, G. Quintin, O. W. Tsang, **D. Gangadharan**, J. Milan and S. Chakraborty, "On buffering with stochastic guarantees in resource-constrained media players", *CODES+ISSS*, 2011, pp. 169-178

[8] B. Raman, A. Nouri, **D. Gangadharan**, M. Bozga, A. Basu, M. Maheshwari, A. Legay, S. Bensalem, S. Chakraborty, "Stochastic modeling and performance analysis of multimedia SoCs", *SAMOS*, 2013, pp. 145-154

[9] A. Weichslgartner, **D. Gangadharan**, S. Wildermann, M. Glaß, J. Teich, "DAARM: design-time application analysis and run-time mapping for predictable execution in many-core systems", *CODES+ISSS*, 2014.

[10] A. Weichslgartner, S. Wildermann, **D. Gangadharan**, M. Glaß, J. Teich, "A Design-Time/Run-Time Application Mapping Methodology for Predictable Execution Time in MPSoCs", *ACM Transactions on Embedded Computing Systems*, 2018.

[11] E. Sousa, **D. Gangadharan**, F. Hannig, J. Teich, "Runtime reconfigurable bus arbitration for concurrent applications on heterogeneous MPSoC architectures", *DSD*, 2014, pp. 74-81.

[12] **D. Gangadharan**, E. Sousa, V. Lari, F. Hannig, J. Teich, "Application-driven reconfiguration of shared resources for timing predictability of MPSoC platforms", *ACSSC*, 2014, pp. 398-403.

[13] D. Golomb, **D. Gangadharan**, S. Chen, O. Sokolsky, I. Lee, "Data Freshness Over-Engineering: Formulation and Results", *ISORC*, 2018 (**Best Paper Award**)

[14] R. Pellizzoni, N. Paryab, M-K. Yoon, S. Bak, S. Mohan, R.B. Bobba, "A generalized model for preventing information leakage in hard real-time systems", *RTAS*, 2015.

[15] **D. Gangadharan**, O. Sokolsky, I. Lee, B. Kim, C.-W. Lin, S. Shiraishi, "Bandwidth Optimal Data/Service Delivery for Connected Vehicles via Edges",*IEEE CLOUD*, 2018 (**Selected as one of the best papers from IEEE CLOUD and invited for a journal publication**).

[16] **D. Gangadharan**, O. Sokolsky, I. Lee, B. Kim, "Multi-Objective Optimization for Data/Service Delivery to Connected Vehicles via Edges", *TREC4CPS*, 2018.

[17] A. Gupta, J. J. Cherukara, **D. Gangadharan**, B. Kim, O. Sokolsky, I. Lee, "E-PODS: A Fast Heuristic for Data/Service Delivery in Vehicular Edge Computing", *IEEE VTC-Spring*, 2021.

[18] A. Gupta, J. J. Cherukara, **D. Gangadharan**, B. Kim, O. Sokolsky, I. Lee, "Global Edge Bandwidth Cost Gradient-based Heuristic for Fast Data Delivery to Connected Vehicles under Vehicle Overlaps", *IEEE VTC-Spring*, 2022.

[19] S. U. Goparaju, Lakshmanan L, A. Navnit, R. Biju, L. Bajaj, **D. Gangadharan**, A. Hussain, "Time Series-based Driving Event Recognition for Two Wheelers", *DATE*, 2023.

[20] S. U. N. Goparaju, SVSLN S. S. Vaddhiparthy, Pradeep C, A. Vattem, **D. Gangadharan**, "Design of an IoT System for Machine Learning Calibrated TDS Measurement in Smart Campus", *IEEE WF-IoT*, 2021.

[21] S. Mante, SVSLN S. S. Vaddhiparthy, R. Muppala, **D. Gangadharan**, A. Hussain, A. Vattem, "A Multi Layer Data Platform Architecture for Smart Cities using oneM2M and IUDX", *IEEE WF-IoT*, 2022.

[22] L. Micconi, **D. Gangadharan**, P. Pop, J. Madsen,"Multi-ASIP platform synthesis for real-time applications", *SIES*, 2013.

[23] **D. Gangadharan**, L. Micconi, P. Pop, J. Madsen,"Multi-ASIP platform synthesis for Event-Triggered applications with cost/performance trade-offs", *RTCSA*, 2013.

[24] L. Jozwiak, M. Lindwer, R. Corvino, P. Meloni, L. Micconi, J. Madsen, E. Diken, **D. Gangadharan**, R. Jordans, S. Pomata, P. Pop, G. Tuveri, L. Raffo, G. Notarangelo,"ASAM: Automatic architecture synthesis and application mapping",*Microprocessors and Microsystems*, 2013.

[25] S. U. Goparaju, K. Pothalaraju, S. Dullur, A. Jain, **D. Gangadharan**."Time-Series based Fall Detection in Two-Wheelers",*IEEE VTC-Fall*, 2023.

[26] S. U. Goparaju, R. Biju, P. Mukkiri,Bhavana MC, **D. Gangadharan**, B. Mandal, Pradeep C, "Optimization and Performance Evaluation of Hybrid Deep Learning Models for Traffic Flow Prediction",*IEEE VTC-Spring*, 2023.

[27] SVSLN S. S. Vaddhiparthy, J. J. Cherukara, **Deepak Gangadharan**, B. Kim, "Collision-Aware Data Delivery Framework for Connected Vehicles via Edges",*IEEE VTC-Fall*, 2023.

[28] J. J. Cherukara, SVSLN S. S. Vaddhiparthy, **Deepak Gangadharan**, B. Kim, "Dynamic Data Delivery Framework for Connected Vehicles via Edge Nodes with Variable Routes", *IEEE VTC-Fall*, 2023.

[29] VJS Pranavasri, L. Francis, U. Mogadali, G. Pal, SVSLN S. S. Vaddhiparthy, A. Vattem, K. Vaidhyanathan, **D. Gangadharan**, "Scalable and Interoperable Distributed Architecture for IoT in Smart Cities", *IEEE WF-IoT*, 2023.