RESEARCH STATEMENT

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 connected vehicles* 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. There are common performance objectives considered in both an IoT system and a RTE system such as end-to-end delay, energy consumption, etc. However, the factors affecting predictability of these performance objectives in IoT systems are many more in comparison to the RTE systems. This poses the requirement for developing robust and efficient analysis methods taking into consideration specific contexts of the two systems in order to analyze the bounds on the performance objectives. In general, I envision to work towards proposing novel design and analysis techniques for *timing predictability*, *energy/thermal efficiency*, *system resilience* and *optimal resource usage.*

1 Prior Work

1.1 Performance Analysis and Design of RTE Systems

I have 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]. Under target quality constraints of the multimedia stream, these techniques were proposed to derive the buffer sizes [1, 2], processing requirements [3] and schedule of drop frames under thermal constraints [4]. Further, I have also developed analytical techniques based on Stochastic Network Calculus to derive quality-of-service (QoS) quantified as probabilistic bounds on performance for multimedia applications running on MPSoC platforms [7, 8]. In addition, I proposed techniques to perform fast hybrid simulation [5, 6] for performance analysis of multimedia MPSoC platforms, whereby the complexity of simulating the execution of video decoders on MPSoC platforms was reduced using the concept of representative video set [5] and by modeling the workload requirement of certain decoding tasks that are computationally intensive [6].

As a PostDoctoral Researcher, I have worked on several problems delving into proposing new design techniques for better timing predictability of executing concurrent applications on MPSoC platforms. 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].

1.2 Analysis and Design of Edge Computing-Based Automotive IoT systems

Delivery of data/services to vehicles via the edge while the vehicle is on the move requires allocation of adequate amount of memory, computation and communication resources on the edge nodes. In my previous work [15], I formulated this data/service delivery problem as an optimization problem, which minimizes the system wide total bandwidth cost of the edge nodes. I also studied the effect of variation of different traffic flow model parameters on the bandwidth cost. 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. However, the optimization frameworks only dealt with the scenario where the vehicles followed a fixed pre-specified route as in the case of a delivery truck fleet belonging to a company, where the routes are assigned beforehand to the truck drivers.

2 Future Research Plan

2.1 Performance Analysis and Design of RTE Systems

Under this theme, I will be working on problems in scheduling and analysis of RTE systems delving into aspects such as *data freshness*, *security over shared resources*, *mixed criticality* and *multimode system operation*, while ensuring timeliness, energy/thermal efficiency and resilience objectives. All these themes will be primarily explored in the context of RTE systems relevant to automotive industry.

One of my recent papers [13] (won the **best paper award**), which was co-authored with a graduate student I mentored, proposed a method to determine the periods of tasks in chains of arbitrary length while satisfying end-to-end freshness constraints with only few assumptions regarding the scheduling algorithm used. I plan to continue this line of research by proposing a data freshness aware scheduling technique for energy/temperature minimization. Current RTE systems are susceptible to faulty execution due to external attacks or faults in the underlying platform. Such a behaviour in the RTE system can adversely affect the freshness bound of task chains. Therefore I would like to explore the effect of faults on data freshness and formulate conditions that will minimize the effect of faults on data freshness.

Security-aware scheduling [14] is an interesting and important research direction, which has many unanswered problems. Within this direction, I would like to propose analysis and scheduling techniques considering dynamic priority scheduling strategies, which have not been considered yet. This problem is even more interesting in a multimode context because the effects on security critical tasks due to low security tasks are enhanced when there are multiple modes of operation. Further, my research will also span the dimension of analysis and scheduling strategies for mixed criticality multimode embedded systems, which is an area I have worked on lately.

2.2 Analysis and Design of Edge Computing-Based Automotive IoT systems

The lines of research that I propose to pursue in future within the area of edge computing-based automotive IoT system broadly fall under the following themes

2.1.1. Handling dynamic route changes

One solution that I would like to explore in this context is to determine the feasibility of a runtime algorithm that will make decisions to transfer the data/service to the appropriate edges when the vehicle changes the route.

2.1.2. Addressing scalability issues

I plan to address this problem using two approaches. The first approach will be to develop a low complexity heuristic algorithm, which can be run online and will deliver data/service to the vehicles such that the solution is close to optimal. In the second approach, I envisage to develop a partitioned optimization approach with local optimizations and a global heuristic handling global data adjustments.

2.1.3. Performance analysis of edge computing for deep learning algorithms

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. Vehicle Platoon Control is one such connected vehicle application which can greatly benefit from DL based edge analytics whereby vehicles can learn a shared prediction model of the environment in which they are moving. This can reduce the amount of data generated by sharing mutual information and as a result reducing the burden on communication and computation. 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.

2.1.4. Performance analysis of concurrently executing safety critical automotive task and deep learning based inference task on a edge device

Currently, there is a lot 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. 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. This problem can also be explored by considering computing on more than one edge device. The relevant question to answer in this case would be to determine the minimal number of edge nodes required to execute the DL inference and safety critical tasks such that a target inference accuracy and timeliness properties are respectively satisfied.

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.