IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-7, Special Issue-16, May 2019 E-ISSN: 2347-2693

A Survey on Performance Optimization of Cache Memory in the Individual Nodes of Wireless Sensor Networks

Amulya V[1*], Mohan K G[2], Ramesh Babu H. S[3]

[1]Department of Computer Science, Sai Vidya College of Engineering, Bangalore, India
[2]Department of Computer Science, Presidency University, Bangalore, India
[3]Department of Computer Science, Sai Vidya College of Engineering, Bangalore, India

Corresponding Author: Amulyav777@gmail.com

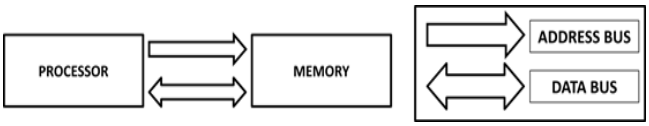
DOI: https://doi.org/10.26438/ijcse/v7si16.7580 | Available online at: www.ijcseonline.org

***Abstract*—** The wireless networks are constrained networks with limited battery backup and memory. The nodes in the Wireless Sensor Network use memory buffers to keep track of the sequence number of the Transport layer segments. This helps to resend the packets during the time of packet loss. The fast retransmission of the lost packets is done by nodes of Wireless Sensor Network with the help of fastest form of memory called as cache memory. This helps to achieve reliability. Understanding the working of cache memory in fulfilling such a great responsibility is a challenge. The survey has been conducted to understand the different types of processors and memories that could be used in the nodes of wireless sensor networks. An overview of optimization methods on cache memory and cache mapping mechanisms to improve the performance of the cache are also studied.

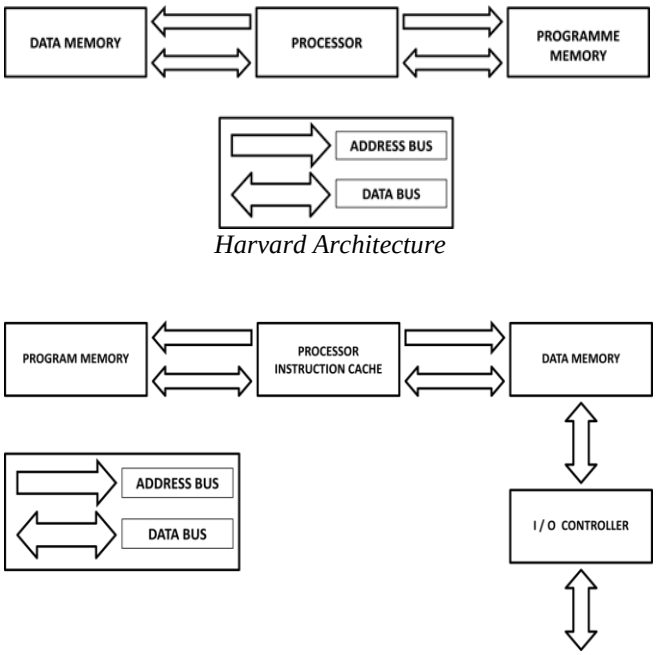
Keywords— *Wireless Sensor Networks (WSN), Cache memory, performance optimization, survey*

I. INTRODUCTION

The wireless networks are constrained networks with limited battery backup and memory. They have the advantages of connecting multiple devices to the same internet connection. They allow mobility of different mobile devices like laptops, mobile phones and tablets to move around in the network area spontaneously and still maintain the connection to the internet. Unfortunately, these networks suffer high chance of packet loss not only due of lossy channels but also due to congestion in the underlying network. To address such an issue, researchers make use of reconfigurable cache at the intermediate nodes. The Nodes of sensor network can be designed using Von Neumann architecture, Harvard architecture or Super Harvard architecture. Figure 1 shows the architecture of the wireless sensor network nodes. Each node in a Wireless Sensor Networks has a Controller, Communication devices, Sensors or actuators, power supply and memory [1,4]. The Figure 2 shows the structure of a wireless sensor node.



Von Neumann Architecture



Harvard Architecture

Super Harvard Architecture

Figure 1. Architectures of Wireless Sensor Network Nodes

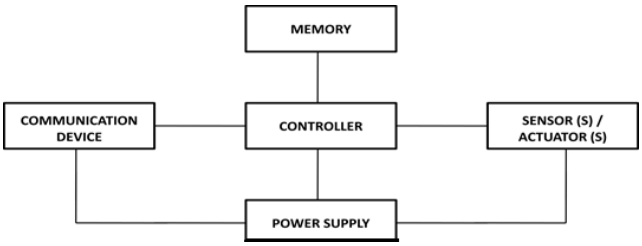


Figure 2. Structure of Wireless Sensor Network Nodes

The cache memory in sensor nodes helps in speeding up the period of time between the Random-Access Memory and the CPU. Pre-fetching of the required data from the main memory for the processor, prior to the processor requests is achieved using cache memory. The locality of reference [2] is the main principal behind data pre-fetching.

The major constraints during the design of any cache memory are the performance, power consumption and the area available as a cache. With these constraints in mind, the cache researchers came up with the three cache organizations called direct mapped cache, associative cache and set associative cache [3]. This survey is proposing to adopt the best possible mapping technique for a sensor node.

II. RELATED WORK

Multiple works have been deployed to increase the performance of the WSNs [5][6]. Some work mainly concentrates on the protocols which utilizes the cache memory to avoid congestion. One such work is implementing the Distributed Transport protocol for Sensor Networks (DTSN). The DTSN is implemented with fixed transmission window and without congestion control mechanism for testing the congestion. The works conclude that the optimum usage of cache is achieved during congestion avoidance.

Adam Dunkels, etal [7] proposed different Mechanisms like Distributed TCP Caching (DTC) for WSNs that have been deployed to increase the reliability of the network. Segment caching and local retransmission using cache are the major ideas in DTC to sidestep end to end retransmissions. This work helps to reduce power consumption by TCP sender & receiver nodes and increase the throughput of the network by caching the segment with highest sequence number at the nodes closer to the receiver. The work Guarantee the throughput of network but not the cache.

Works on protocols like End-to-end Reliable and Congestion aware Transport layer Protocol (ERCTP) [8] for heterogeneous WSNs are surveyed in this work. The protocol developed is divided in to two modules namely congestion control module and Reliable module. The congestion control module is further divided in to congestion detection module, congestion notification module, rate adjustment modules. This paper concludes that, by introducing distributed memory caching, helps to aware the congestion by congestion aware module, and reliability by reliable module. Performance optimization techniques for optimization of WSNs such as Cross-layer caching [8] were implemented to understand the transport layer NACK repair and MAC layer adaptive retransmission processes to achieve reliability. The work also concludes that NACK and adaptive retransmission enriches the advantage of intermediate caching. Works related to reducing the overall energy ingesting and the file transfer extent were conducted by improving DTCP Caching for WSNs [9]. Each node will cache the TCP segments. The acknowledged segments are unlocked, and unacknowledged segments are locked. This helps in retransmission and in turn helps in reliability. The author states, that these types of works help in understanding TCP window size and TCP caching performance.

Different surveys of transport protocol for WSNs [10] have been conducted to help to understand the different transport protocols that can be used in WSN. But, the study discusses only about the retransmissions of the lost data during transmission but not specifies how it is to be retransmitted. The study also does not mention the use of cache memory in the nodes.

Protocols like Reliable Multi-Segment Transport (RMST) [11] were introduced to propose a diffusion architecture where the sink subscribes to the interest. The interest will then publish the type of data it is interested in the form of attribute & value pairs. All other nodes having the data will revert. Cache is used in 2 modes, the mode of caching and a mode of non-caching. In mode of caching, one major node maintains a cache locally & all other nodes make use of the major cache. In mode of non-caching every sources and sinks sustain their own cache. Cache entries have fragment map (actual data) & hole map (timer). The work fails to specify which cache mapping techniques to be used to achieve maximum usage of both cache modes in order to achieve reliability.

Pump- slowly, fetch quickly (PSFQ) [12] is a Reliable transport protocol for sensor networks, that was introduced to help data distribution from source at very slow rate and collection of lost data from the neighbouring nodes at highly aggressive rates. The pump function of PSFQ likely performs controlled flooding and needs every transitional node to have a data cache to be used during local loss recovery and during the time of in-sequence data delivery. To improve the reliability of a WSN Multi-path-based Distributed TCP Caching (MDTC) [13] was introduced. MDTC implemented path redundancy and local retransmission. When each node receives the data, it caches the data and sets timer for the incoming data. The nodes wait for the ACK of the packet. If the ACK is not received in the incoming data timer, node's

cache is locked and timer for retransmission is started. When the timer for retransmission elapses, the cached packet in the nodes are used for retransmission and the transmission timer is set again.

Works on cache coherence within the networks [14] were implemented to study the read and write requests of each node in the network. The work does not specify which cache mapping techniques to be used during read and writing in to the cache.

Different MAC protocols for energy efficient Wireless Sensor Networks [15] were developed to make the network energy efficient by involving the nodes in periodic listen and sleep modes to reduce the energy consumption. But, the work fails to specify the contribution of cache memory in saving energy during retransmissions of lost or corrupted data.

In all the literatures cited above, most of the authors did not specify which cache mapping and optimization techniques to be used during reading and writing of data in to the cache.

III. METHODOLOGY

To optimize the performance of cache memory in the nodes of wireless sensor networks, the following points are to be considered.

- Miss rate reduction
- Miss penalty reduction
- Reduction in the time to hit in the cache

A. Miss Rate Reduction

While attempting to access data from the cache, if the process is unsuccessful, main memory access is required to retrieve the data and long latency in cache access is encountered. This phenomenon is known as cache miss. The survey studies the three different types of cache misses [16] known as Compulsory miss which is also known as Cold miss, Conflict miss and Capacity miss. The Cache misses encountered during the first-time access are called as compulsory misses.
Conflict miss is a cache miss that occur when we require a data that was available in the cache some time back but erased from the cache currently. Conflict miss can be measured by considering the difference between the numbers of miss of two cache memories with same cache size and block size, but one with limited and other with full associativity. The Capacity miss is a cache miss that occurs because of the availability of less size in the cache memory irrespective of their mapping techniques. Capacity misses are the cache misses that are very difficult to identify and rectify. The cache misses that still appear after taking the difference between compulsory and conflict misses are called as capacity misses.

B. Miss penalty Reduction

The average memory access time of a processor decides the performance of a processor and can be given as in equation (1),

$$AT=HT+MR*MP \quad (1)$$

Where, AT is the average access time, HT is a Hit time which can be defined as the time required to deliver a block of data from cache to processor, MR is Miss rate and can be defined as the cache misses per cache references and MP is the miss penalty which can be defined as additional time that is required for a processor due to cache miss. Miss penalty directly affects the average access time.

C. Reduction in the time to hit in the cache

The data requested are not available in the cache until address translations are performed. Hence, there is need to avoid address translations while indexing the cache in order to reduce hit time. For this purpose, set index are taken from the page offset portion of the virtual address.
Address translation can also be avoided by indexing the cache directly by virtual address.

To reduce the cache miss rates, different optimization methods or techniques are used. The compulsory misses are reduced by the Data prefetching technique. Data prefetching helps the microprocessor to request for the data whose computation is not already started. Prefetching instructions can either be hardware or a software. An example of hardware prefetching is Sequential prefetching.

Prefetching of the data by cache memory will be successful if the data is predicted. Data prediction can be done either by hardware or by the software instruction specially designed for data prefetching. Data prefetching will also be successful when there exists enough space in the cache memory to handle the previously retrieved data and the previously fetched data that are active.

Compulsory misses are condensed by snowballing the block size of the cache to certain extent to maintain spatial locality. To certain extent, increasing the associativity can reduce the conflict miss. The conflict miss can also be reduced by changing the size of the arrays used as data, changing the starting address of the variables, array transpositions and combining multiple arrays. Array padding [18, 10 and 20], Intra array padding, array merging or group and transpose [21], Data copying are some of the techniques to reduce the conflict miss. All of these techniques are termed as Optimization of data layouts, which helps not only in reducing the conflict miss but also the false sharing [17].
Capacity miss can be reduced by using loop blocking or loop tiling techniques. [22, 23, 24].
To reduce the miss penalty there are four following techniques,

- Multi-level Caches

- Critical word first and early restart
- Properties to read misses over write misses
- Victim caches

A. Multi Level Caches

Multilevel caching is a technique used to avoid miss penalty by introducing another level of cache between the actual cache and the main memory. The technique requires extra hardware support.

B. Critical word first and early Restart

Critical word first technique first requests for the word that was missed in the cache and sends it to CPU as soon as it is available. The technique is also called as wrapped fetch. The technique is impatient since it does not allow complete word to be loaded.

Early restart helps to fetch the data in the normal order but when the requested data is fetched it sends the data with highest priority to the processor.

C. Properties to read misses over write misses

Read misses are the cache misses that occur when an attempt is made to read the data from the cache before it was written in to the cache. To avoid read misses a write-through cache with write buffer of enough size can be used.

D. Vicctim Caches

The victim cache technique uses fully associative cache memory of a tiny size to remember the data that was discarded previously. The idea behind the approach is that the data discarded may be needed in the near future.

IV. OBJECTIVE AND DISCUSSION

The critical factor of a node in the WSN is its cache memory. The cache is designed to overcome the speed gap during instruction execution and hence it is important to study the performance measurement and metrics of cache memory during the design of the WSN nodes.

The cache misses and the cache hits play a vital role in determining the performance of the cache memory. The cache misses and hits help to determine the Average Access Time, which is one of the most important performances metric.

Other than Average Access Time, performance of the cache also depends on cache size and cache mapping techniques. Hence in the proposed survey following objectives have been considered.

- To propose best suited cache mapping and optimization mechanisms to reduce conflict misses and miss penalties in the nodes of WSN.
- To propose an optimal size for a cache memory to reduce capacity misses, a factor of miss rate in the nodes of WSN
- To propose energy efficient and reliable cache performance WSN.

V. CONCLUSION AND FUTURE SCOPE

The wireless networks are constrained networks with limited battery backup and memory. The nodes in the WSN use memory buffers to keep track of the sequence number of the Transport layer segments. This helps to resend the packets during the time of packet loss. The fast retransmission of the lost packets is done by nodes of WSN with the help of fastest form of memory called as cache memory. This helps to achieve reliability. Understanding the working of cache memory in fulfilling such a great responsibility is a challenge. The structure of wireless sensor node consists of a Controller, Communication devices, Sensors or actuators, power supply and memory. The controllers that can be used in the wireless sensor nodes can be Texas Instruments MSP430 and Atmel ATMega 128L. Texas Instruments MSP430 is a 16-bit, RISC core processor with the clock cycle of up to 4 MHz and available in different versions with 2-10 kb RAM. Texas Instruments MSP430 is provided with several DACs and the price starts at 0.49 US$. Atmel ATMega 128L is a 8-bit controller, with larger memory than MSP430 but, it is slower than MSP430.

The MSP430F15x/16x/161x are microcontroller configurations that has dual 12-bit Digital to Analog converter, dual in-built 16-bit timers, a high speed 12-bit Analog to Digital converter, at most two universal serial synchronous or asynchronous communication interfaces (USART), I2C, 48 Input or Output pins and Direct memory access (DMA). The process in which the sub system of hardware of a computing device gets access to the random-access memory without the intervention of central processing unit (CPU) is known as Direct Memory Access. MSP430 series provides extended addressing of Random-Access Memory for applications which are memory-intensive and applications with the requirements of large C-stack.

It is observed from the survey, that most of the WSN nodes use the controllers of Atmel family. There are two spaces of main memories in Atmel Atmega 128L, the data and the program memory spaces. It also consists of EEPROM used for data storage, contains in system on-chip flash memory of 128kb and reprogrammable for program storage. The flash program memory is further divided in to two parts and are called as Boot program section and application program section. Atmega 128L also contains a SRAM which acts as a cache memory and has a size of 4kb.

Understanding the working of SRAM of Atmega 128L as a cache memory and improving the performance of SRAM as a cache in order to improve the reliability is a challenge. Summarizing the above paragraph, the main problems involved in improvising the performance of WSN are,

- To determine most favorable cache size to account for the cache hits / misses to optimize the performance.
- To adopt best mapping and optimizing techniques to read and write the data into cache for better results in performance of the cache and entire network.
- To analyze network reliability and energy efficiency based on the impact of change in cache size and mapping technique.

The Power law of cache misses helps to test the performance of the cache memory by increasing the size of the cache memory and testing the number of cache miss. This testing technique requires having a reference of number of cache miss of any one particular cache size. Hence, in the further research the output of Atmel Atmega 128L microcontroller will be taken as the reference cache miss.

The performance of the cache can also be tested using a technique called as Stack distance profiling, in which cache miss is measured by implementing the different associativity or the cache mapping techniques on the cache. The proposed work uses this technique to Compare different cache mapping techniques currently in use and obtain the best cache mapping technique for enhancing the performance of WSN. With the results of testing the performance of the cache of WSN node, the best cache size and mapping technique of the cache is determined and can be tested on different protocols like, DTCP, ERTCP, RMST, PSFQ and MTDC which could help in increasing the reliability of the WSN.

In this survey an effort is made to investigate the performance of the WSN by understanding the different types of cache mapping techniques on the cache memory of different size of an individual node in a WSN. The work also helps to adopt the best suited mapping mechanisms and optimal size for the cache memory of the nodes to reduce conflict misses and capacity misses in the cache. The work helps to optimize the performance of the cache and also the entire WSN. Increasing the performance of cache memory in the individual nodes of WSNs will in turn decrease the delay in resending the lost packets in the WSN. Thus, the performance of the entire WSN is increased.

Proposed work helps in increase the speed of packet retransmission by optimizing the performance of the cache and thus we can increase the reliability of the network. This work can also help to save energy - The most important resource of the WSNs.

Wireless sensor networks are the constrained networks with limited resources like battery and memory. Multiple works have been deployed in order to save energy and increase the performance of the WSN in terms of reliability, security and scalability.

Proposed work has surveyed different literatures used to increase the performance of the WSN. Most of the researchers propose different protocols that increase the reliability of the network by caching the data in the cached memory and using the cached data for retransmission during the situation of congestion.

However most of these researches does not show the structure, implementation, size and mapping mechanisms of the cache memory used to store the data used for retransmission.

The survey and hence the future work help in understanding the structure, implementation, size, mapping mechanisms and different optimization techniques of the cache memory used to store the data which is used for retransmission and thus helps to optimize the performance of individual nodes in the WSNs in terms of reliability and energy efficiency.

ACKNOWLEDGMENT

This work was supported by Research centre-Sai Vidya institute of technology, Bangalore, India and Presidency University, Bangalore, India.

REFERENCES

[1] Waltenegus Dargie and Christian Poellabauer, **"***Fundamentals of Wireless sensor theory and practice*"

[2] Safaa S. Omran and Ibrahim A. Amory, "*Design of Two-Dimensional Reconfigurable Cache memory using FPGA*", IEEE 10.1109/ICEDSA.2016.7818502, 2016,

[3] Santana Gil, A. D. Benavides Benitez, J.I., Hernandez Calviño, M. , Herruzo Gómez, E , "*Reconfigurable Cache implemented on an FPGA*", International Conference on Reconfigurable Computing, IEEE 10.1109/ReConFig.2010.26, 2010

[4] Dinesh Kumar Gupta, "*A Review on Wireless Sensor Networks*", Inter-national Conference on Recent Trends in Applied Sciences with Engineering Applications, ISSN 2224- 610X, Vol.3, No.1, IEEE 10.1109/ICAL.2012.6308240, 2013

[5] Melchizedek I. Alipio and Nestor Michael C. Tiglao," Analysis *of Cache-based Transport Protocol at Congestion in Wireless Sensor Networks*", International Conference on Information Networking (ICOIN)**,** IEEE, 10.1109/ICOIN.2017.7899459, 2017

[6] Bruno Marchil, Antonio Grilo, Mario Nunes, "*DTSN: Distributed Transport for Sensor Networks*", 12th IEEE Symposium on Computers and Communications, IEEE, 1-4244-1521-7/0, 2007

[7] Adam Dunkels, Juan Alonso, Thiemo Voigt, "*Distributed TCP caching for wireless sensor networks*", SICS Technical Report T2004:06 ISSN 1100-3154 ISRN: SICS-T–2004/06- SE, 2004.

[8] Atif Sharif, Vidyasagar M. Potdar, A. J. D Rathnayaka, "*ERCTP:*

End-to-end Reliable and Congestion aware Transport layer Protocol for heterogeneous WSN", Scalable Computing: Practice and Experience, vol. 11, no. 4, 2010.

[9] Ahmed Ayadi, Patrick Maill´e, and David," *Improving Distributed TCP Caching for Wireless Sensor Networks*", The 9th IFIP Annual Mediterranean AdHoc Networking Workshop (Med-Hoc-Net)**,** IEEE 10.1109/MEDHOCNET.2010.5546858, 2010.

[10] Su Liu1, Yan Tang*, Yonghua Liu2, "*A survey of transport protocol for wireless sensor networks*", 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet), IEEE 10.1109/CECNet.2012.6202037, 2012

[11] Fred Stann, John Heidemann, "*RMST: reliable data transport in sensor networks*", Proceedings of the First IEEE International Workshop on Sensor Network Protocols and Applications, IEEE 10.1109/SNPA.2003.1203361, 2003

[12] Chieh-Yih Wan, Andrew T. Campbell, Member, IEEE, and Lakshman Krishnamurthy, "*Pump-slowly, fetch quickly (PSFQ): a reliable transport protocol for sensor networks*", IEEE Journal on Selected Areas in Communications, vol. 23, no. 4, IEEE 10.1109/JSAC.2005.843554 , 2005

[13] Yuhua Liu and Hao Huang, "*Multi-path-based Distributed TCP Caching for Wireless Sensor Networks*", Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD)**,** IEEE 10.1109/SNPD.2007.297, 2007

[14] Noel Eisley, Li-ShiuanPeh, and _Li Shang, " *In- Network Cache Coherence*", IEEE Computer Architecture Letters , Vol. 5, IEEE Computer Architecture Letters, 2006

[15] Wei Ye, John Heidemann, Deborah Estrin , "*An Energy-Efficient MAC Protocol for Wireless Sensor Networks*", Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies**,** IEEE 10.1109/INFCOM.2002.1019408, 2002

[16] Markus Kowarschik and Christian Wei, "*An Overview of Cache Optimization Techniquesand Cache Aware Numerical Algorithms*", Deutsche Forschungsgemeinschaft (German Science Foundation), projects Ru 422/7-1,2,3.

[17] J.L. Hennessy and D.A. Patterson, "*Computer Architecture: A Quantitative Approach*" Morgan Kaufmann Publisher, Inc., San Francisco, California, USA, second edition, 1996.

[18] M.S. Lam, E.E. Rothberg, and M.E. Wolf, "*The Cache Performance and Optimizations of Blocked Algorithms*" In Proc. of the Fourth Int. Conference on Architectural Support for Programming Languages and Operating Systems, pp 63-74,Palo Alto, California, USA, 1991.

[19] J. Torrellas, M. Lam, and J. Hennessy, "*Shared Data Placement Optimizations to Reduce Multiprocessor Cache Miss Rates*", In Proc. of the Int. Conference on Parallel Processing, volume 2, pp 266-270, Pennsylvania, USA, 1990.

[20] G. Rivera and C.-W. Tseng, "*Data Transformations for Eliminating Conflict Misses*" In Proc. of the ACM SIGPLAN Conference on Programming Language Design and Implementation, Montreal, Canada, 1998.

[21] T. Jeremiassen and S. Eggers, "*Reducing False Sharing on Shared Memory Multiprocessors through Compile Time Data Transformations*" In Proc. of the Fifth ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming, pp 179-188, Santa Barbara, California, USA, 1995.

[22] R. Allen and K. Kennedy, "*Optimizing Compilers for Modern Architectures*" Morgan Kaufmann Publishers, San Francisco, California, USA, 2001.

[23] Y. Song and Z. Li, "*New Tiling Techniques to Improve Cache Temporal Locality*", In Proc. of the ACM SIGPLAN Conference on Programming Language Design and Implementation, pp 215-228, Atlanta, Georgia, USA, 1999.

[24] M.E. Wolf and M.S. Lam "*A Data Locality Optimizing Algorithm*" In Proc. of theSIGPLAN'91 Symposium on Programming Language Design and Implementation, volume 26 of SIGPLAN Notices, pp 33-44, Toronto, Canada, 1991.

Authors Profile

Mrs. Amulya V pursed Bachelor of Engineering from Visveswaraya technological University, Belgaum, India in 2011 and Master of Technology from Visveswaraya technological University in year 2014. she is currently pursuing Ph.D. She is a former Assistant Professor in Department of Computer Sciences, East west College of Engineering Bangalore and currently working as a data analyst in Redington Gulf, India Branch.

Dr. K.G Mohan, has 33 years of Teaching and Administrative experience working in different cadre like Assistant Professor, Associative Professor, Professor,Head of the Department, Dean-R&D, Dean-Academics and Principal. His reasearch area includes Energy Efficient Processor subsystems, High perfromance computing and WSN. He is a Former member of BoS, CSE and BoE, CSE in Visvesvaraiah Technological University, Belgaum, and Academic Council member of B.V.B.C.E.T, Hubli, India

Dr. Ramesh Babu H S, , has 25 years of Teaching and Administrative experience working in different cadre like Assistant Professor, Associative Professor, Professor,Head of the Department, Dean-R&D, Dean-Academics and Principal. He has worked as a free-lance Consultant for Tech Mahindra. He is a certified corporate trainer for Programming techniques, C and Software Engineering courses.

