

UBAID ULLAH HAFEEZ

(+1) 631-215-8668 | uhafeez@cs.stonybrook.edu

Education

Stony Brook University (SUNY), Stony Brook, NY

Anticipated graduation: 2021

Ph.D. in **Computer Science**

CGPA:3.85/4.00

Advisor: [Professor Anshul Gandhi, Ph.D.](#)

Lahore University of Management Sciences (LUMS), Lahore, Pakistan

Bachelor of Science (Hons.) [**Major: Computer Science**]

Aug '16

Course Work: Networks, Network Security, Topics in Internet Research, Big Data Analytics, Applied Probability, Multicore Computing, Machine Learning, Data Science, Compiler Design, Computer Vision

Publications

Ubaid Ullah Hafeez, Muhammad Wajahat, Anshul Gandhi - "***EIMem: Towards and Elastic Memcached System***". In *Proceedings of IEEE 38th International Conference on Distributed Computing Systems (ICDCS) 2018*, VIENNA, Austria, July 2018 (**Best Student Paper Award**)

Ubaid Ullah Hafeez, Deepthi Male, Sharath Kumar Naeni, Muhammad Wajahat, Anshul Gandhi - "***Realising an Elastic Memcached via Cached Data Migration***". In *Proceedings of Middle Ware 2017*, LAS VEGAS, NV, USA, December 2017 (Poster)

Ubaid Ullah Hafeez, Aqsa Kashaf, Qurat ul ain Bajwa, Aisha Mushtaq, Hassan Mujtaba Zaidi, Ihsan Ayyub Qazi, Zartash Afzal Uzmi - "***Mitigating Datacenter Incast Congestion Using RTO Randomization***". In *Proceedings of IEEE GLOBECOM 2015*, SAN DIEGO, CA, USA, December 2015

Awards and Honors

- ICDCS Best Student Paper Award 2018
- ICDCS Student Travel Award 2018
- Stony Brook Graduate Fellowship Award for three consecutive years, 2016-19
- Graduation with distinction, [LUMS](#) Dean's Honor List for three consecutive years 2013-16
- NOP Scholarship Award for four years, 2012-16

Research and Work Experience

Graduate Student Researcher

- Stony Brook University, NY, USA

Jun '17 – Present

Summer Research Intern

- IBM T.J. Watson Research Center, NY, USA

May '18 – Aug '18

Teaching Assistant

- CSE 214: Data Structures and Algorithms at Stony Brook University

Aug '16 – Dec '16

Teaching Assistant

- Design and Analysis of Algorithms at IT University, Lahore, Pakistan

Oct '15 – Jan '16

Research Assistant

- Lahore University of Management Sciences, Lahore, Pakistan

Jun '15 – Nov '15

Teaching Assistant

- CS 200: Introduction to programming at Lahore University of Management Sciences

Aug '14 – Dec '14

Selected Projects

Automated Patch Management in Hybrid Cloud

May '18 – Aug '18

Position: Summer Research Intern at IBM T.J. Watson Research Center

Patching vulnerabilities in online applications is very important as most of the attacks are not zero-day attacks. For applications deployed in hybrid cloud, services deployed in different infrastructures have different procedures for patching. Also, typically there is a small timing window during which applications can be brought down and patched. I developed a framework which automatically applies patches to different components of application deployed in different infrastructures while prioritizing high risk vulnerabilities when all the patches cannot be applied in the given timing window.

TensorFlow Device Placement

Feb '18 – Present

Position: Graduate Student Researcher at Stony Brook University

In the recent past, we have seen a significant increase in size and computational requirements for training and inference with neural networks. Typically, a heterogenous environment with a mixture of devices such as CPUs and GPUs, is used to address these requirements. The decision of partitioning the neural model across devices is closely related to the efficiency and utilization of these devices. This decision is often made by human experts based on simple heuristics or intuitions which sometimes leads to in-efficient utilization of resources. I am currently working on devising an algorithm and framework which can partition the neural net, expressed as a TensorFlow graph, such that all the devices being used are fully utilized at all times to achieve close to optimal performance.

Elastic Memcached (EIMem)

Jun '17 – Dec '17

Position: Graduate Student Researcher at Stony Brook University

Memory caches, such as Memcached, are a critical component of online applications as they help maintain low latencies. However, memory caches are expensive, both in terms of power and operating costs. It is thus important to dynamically scale such caches in response to workload variations. Unfortunately, stateful systems, such as Memcached, are not elastic in nature. The performance loss that follows a scaling action can severely impact latencies. I contributed in developing an elastic Memcached system that mitigates post-scaling performance loss by proactive migration of hot data between cache nodes. Our experimental results on OpenStack, across several workload traces, show that elastic Memcached scales while reducing the post scaling performance degradation by about 90%.

Mitigating Datacenter Incast Congestion Using RTO Randomization

Jun '15 – Aug '15

Position: Research Assistant at Lahore University of Management Sciences

TCP incast congestion happens in many-to-one communication patterns that frequently arise in large-scale datacenter applications such as web search, social networks etc. Incast congestion can severely degrade the performance of applications. To mitigate successive timeouts, I designed algorithms for randomizing TCP retransmission timeout (RTO). These algorithms rely on (a) successive timeouts, (b) explicit knowledge of the level of multiplexing, and (c) the knowledge of flow sizes. Results show that these algorithms improve goodput by 1.5x-11x for up to 64 senders and provide greater improvement for larger number of senders.

Integrity File System (IFS)

Sep '15 – Nov '15

Position: Research Assistant at Lahore University of Management Sciences

Research scientists in computational biology and other data intensive fields typically confront with large data files ranging from hundreds of MBs to a few GBs. Fast validation of these files is often required in scientific computing where integrity of data is of particular concern. To address this use case, I took part in developing a user space file system in FUSE - a user space filesystem infrastructure. To support fast reads and writes while validating the files, I contributed in developing a B+-Tree to store files in the form of chunks. I also worked on parallelizing checksum calculation and writing of data on disk to further reduce overhead of validation.

Content Based Video Retrieval

Sep '15 – Dec '15

Side Project for fun

Implemented a video search engine using regression with both keyword and image-based search. I tested this on a small dataset of episodes of T.V. shows and got 90% accurate results. With the tremendous growth in machine learning in the near past, this idea can be scaled up by using more sophisticated machine learning techniques to support a wide variety of data.

Skills and Abilities

Python, C++, Java, MATLAB, Memcached, TensorFlow, NS (Network Simulator), Android, Bash, PHP, SQL, FUSE