

## Research Interests

### Distributed systems, storage and I/O systems, cloud computing, IoT

The overarching goal of my research is to enable efficient and flexible systems for the growing data demands of modern high-end applications running on existing as well as emerging computing platforms. My current ongoing work focuses on (i) designing efficient systems on resource constrained devices for IoT, (ii) highly-reusable distributed storage systems, and (iii) flash array efficiency optimization.

## Appointments

- 08/2017– **Assistant Professor**, *George Mason University*, Fairfax, VA.  
Department of Computer Science
- 2011–2017 **Research Assistant**, *Virginia Tech*, Blacksburg, VA.  
Department of Computer Science
- 06/2015–12/2015 **Research Intern**, *EMC*, Princeton, NJ.  
Offline flash caching. Mentor: Fred Douglass
- 05/2014–08/2014 **Research Intern**, *IBM Research–Almaden*, San Jose, CA.  
Cloud analytics storage tiering. Mentor: Aayush Gupta
- 05/2013–08/2013 **Research Intern**, *IBM Research–Almaden*, San Jose, CA.  
Load balanced in-memory caching. Mentors: Aayush Gupta, Anna Povolnina

## Education

- 2011–2017 **Virginia Polytechnic Institute and State University (Virginia Tech)**, Blacksburg, VA.  
Ph.D. in Computer Science  
Advisor: Prof. Ali R. Butt
- 2005–2009 **Beijing University of Posts and Telecommunications (BUPT)**, Beijing, China.  
B.Eng. in Computer Science

## Publication

- [ATC '16] **Yue Cheng**, Fred Douglass, Philip Shilane, Michael Trachtman, Grant Wallace, Peter Desnoyers, and Kai Li. *Erasing Belady's Limitations: In Search of Flash Cache Offline Optimality*. In Proceedings of the 2016 USENIX Annual Technical Conference (**ATC'16**), Denver, CO (AR: 47/266 = 17.7%).
- [HotStorage '16] Ali Anwar, **Yue Cheng**, Hai Huang, and Ali R. Butt. *ClusterOn: Building Highly Configurable and Reusable Clustered Data Services using Simple Data Nodes*. In Proceedings of the 8<sup>th</sup> USENIX Workshop on Hot Topics in Storage and File Systems (**HotStorage'16**), Denver, CO (AR: 24/65 = 36.9%).
- [HPDC '16] Ali Anwar, **Yue Cheng**, Aayush Gupta, and Ali R. Butt. *MOS: Workload-aware Elasticity for Cloud Object Stores*. In Proceedings of the 25<sup>th</sup> ACM Symposium on High-Performance Parallel and Distributed Computing (**HPDC'16**), Kyoto, Japan (AR: 20/129 = 15.5%).
- [VarSys '16] Ali Anwar, **Yue Cheng**, and Ali R. Butt. *Towards Managing Variability in the Cloud*. In Proceedings of the 1<sup>st</sup> IEEE International Workshop on Variability in Parallel and Distributed Systems (**VarSys'16**), Chicago, IL.

- [Internet Computing] **Yue Cheng**, M. Safdar Iqbal, Aayush Gupta, and Ali R. Butt. *Provider versus Tenant Pricing Games for Hybrid Object Stores in the Cloud*. In IEEE Internet Computing's special issue on Cloud Storage: May/June 2016.
- [PDSW '15] Ali Anwar, **Yue Cheng**, Aayush Gupta, and Ali R. Butt. *Taming the Cloud Object Stores with MOS*. In Proceedings of the 10<sup>th</sup> ACM Parallel Data Storage Workshop (**PDSW'15**), Austin, TX (AR: 9/25 = 36%).
- [APSys '15] Ali Anwar, **Yue Cheng**, Aayush Gupta, and Ali R. Butt. *Taming the Cloud Object Stores with MOS*. In the 6th ACM SIGOPS Asia-Pacific Workshop on Systems (**APSys'15**), Tokyo, Japan (Poster).
- [HotCloud '15] **Yue Cheng**, M. Safdar Iqbal, Aayush Gupta, and Ali R. Butt. *Pricing Games for Hybrid Object Stores in the Cloud: Provider vs. Tenant*. In Proceedings of the 7<sup>th</sup> USENIX Workshop on Hot Topics in Cloud Computing (**HotCloud'15**), Santa Clara, CA (AR: 21/64 = 32.8%).
- [HPDC '15] **Yue Cheng**, M. Safdar Iqbal, Aayush Gupta, and Ali R. Butt. *CAST: Tiering Storage for Data Analytics in the Cloud*. In Proceedings of the 24<sup>th</sup> ACM Symposium on High-Performance Parallel and Distributed Computing (**HPDC'15**), Portland, Oregon (AR: 19/116 = 16.4%).
- [EuroSys '15] **Yue Cheng**, Aayush Gupta, and Ali R. Butt. *An In-Memory Object Caching Framework with Adaptive Load Balancing*. In Proceedings of the 10<sup>th</sup> ACM European Conference on Computer Systems (**EuroSys'15**), Bordeaux, France (AR: 32/154 = 20.8%).
- [SoCC '13] **Yue Cheng**, Aayush Gupta, Anna Povzner, and Ali R. Butt. *High Performance In-Memory Caching through Flexible Fine-Grained Services*. In Proceedings of 2013 ACM Symposium on Cloud Computing (**SoCC'13**), Santa Clara, CA (Poster).

## Research Experience

- 12/2015–05/2017 **Universal Distributed Storage Management**  
 We argue that for most modern storage applications, the common feature implementation (i.e., distributed management such as replication, consistency, etc.) can be automated and offloaded, so developers can focus on the core application functions. We are designing a framework that aims to reduce the engineering efforts required to develop distributed storage applications. (**HotStorage'16, under submission...**)
- 06/2015–12/2015 **Offline Flash Caching**  
 We explored offline algorithms for flash caching, in terms of both performance and flash lifespan. We designed and implemented a multi-stage heuristic that approximates the offline optimal algorithm. Our investigation provides a useful approximate baseline for evaluating any online algorithm, highlighting the importance of creating new policies for caching compound blocks in flash. (**ATC'16**)
- 03/2015–01/2016 **Re-architecting Object Stores**  
 Existing object stores' monolithic setup results in all different types of applications experiencing the same service level. Extant object store segmentation policies support storage partitioning for serving heterogeneous workloads. However, this is not comprehensive enough and is not capable of supporting dynamically changing workloads. We designed and implemented a micro object storage architecture for the cloud with independently configured microstores each tuned dynamically for a particular type of workload. (**ApSys'15, PDSW'15, HPDC'16**)
- 01/2015–05/2015 **Pricing Games for Cloud Object Stores**  
 The CAST work innovates a way to ensure cost-effectiveness for both cloud service provider and tenants, the goals of whom are essentially contradictory. We designed a tiered object store for the cloud, which comprises both fast and slow storage devices. Simulations show that our approach allows both a service provider and its tenants to engage in a pricing game, which our results show yields a win-win situation. (**HotCloud'15, Internet Computing'16**)

- 08/2014–01/2015 **Tiering Storage for Cloud Data Analytics**  
Comprehensive studies are lacked to guide cloud service users to make the best choice in deploying storage resources for data analytics in the public clouds. We designed and implemented a tiering framework, CAST, and evaluate it on a 400-core cloud cluster using production traces from Facebook. Our tiering mechanism can effectively achieve tenants' goals. (HPDC'15)
- 05/2013–10/2014 **Versatile, Load Balanced In-Memory Caching**  
High performance in-memory object caches suffer from drastic performance degradation due to load imbalance under skewed workloads. We designed and implemented a load balancing approach that involves multiple local and globally coordinated phases: the goal is to decide the stage that is most economical yet effective in proceeding. Experimental evaluation on a 20-node cluster on Amazon EC2 demonstrates the efficacy of our design. (SoCC'13, EuroSys'15)

## Honors & Awards

- 2015 Student Travel Grant: USENIX ATC'15, ACM HPDC'15, EuroSys'15
- 2014 Student Travel Grant: USENIX OSDI'14, FAST'14
- Spring 2014 Pratt Fellowship awarded by CS@VT, *Best TA Award*
- 2013 Student Scholarship: ACM SoCC'13
- 2012 Student Travel Grant: USENIX OSDI'12
- 2006–2009 University Scholarship awarded by BUPT, China
- 2006 Third Prize in the 23<sup>rd</sup> Beijing Physics Competition for College Students

## Proposal Writing Experience

- NFF *Designing a Flexible High-Performance Distributed Storage Framework*, 2017
- CSR: Large *VarSys: Managing Variability in High-Performance Computing Systems*, 2016 (Funded)
- XPS: Small *A Scalable Programming, Management, and Update Framework for the Internet of Things*, 2016 (Funded)
- IBM Research *IBM Faculty Fellowship*, 2015 (Funded)
- CRI *A Microserver-Equipped Hybrid Architecture Testbed*, 2015
- NFF *Mizan: An In-Memory Object Caching Framework with Adaptive Load Balancing*, 2015 (Funded)
- CSR: Medium *Pythia: An Application Analysis and Online Modeling Based Prediction Framework for Scalable Resource Management*, 2014 (Funded)

## Teaching Experience

- Spring 2017 **Guest lecturer**, CS3214 *Intro. to Computer Systems*, CS@VT.  
Led lectures of two sessions
- Spring 2016 **Guest lecturer**, CS6204 *Cloud Computing*, CS@VT.
- Spring 2015 **Teaching assistant**, CS2506 *Computer Organization II*, CS@VT.  
Held office hours; graded homework/projects/exams; proctored exams
- Fall 2014 **Guest lecturer**, CS5204 *Operating Systems*, CS@VT.  
Led discussions
- Spring'12–Fall'14 **Teaching assistant**, CS2505 *Computer Organization I*, CS@VT.  
Held office hours; graded homework/projects/exams; proctored exams
- Spring 2013 **Teaching assistant**, CS5504 *Computer Architecture (grad-level)*, CS@VT.  
Designed presentation schedule + mid-term/final exams; graded projects/exams; proctored exam

- Fall 2011 **Teaching assistant**, *CS3114 Data Structures and Algorithms*, CS@VT.  
Held office hours; graded homework/projects
- Spring 2011 **Teaching assistant**, *CS2073 Computer Programming with Eng. Application*, CS@UTSA.  
Held lab/office hours; graded homework
- Fall'09–Fall'10 **Teaching assistant**, *CS3733 Operating Systems*, CS@UTSA.  
Held lab/office hours; graded homework

---

## Mentoring Experience

- Fall 2016 **Term project mentor**, *CS5204 Operating Systems*, CS@VT.  
Guided 6 students on OS course projects
- Spring 2016 **Research project mentor**, CS@VT.  
Guided 2 students on distributed system projects

---

## Talks

- **Erasing Belady's Limitations: In Search of Flash Cache Offline Optimality**  
[06/2016] USENIX ATC'16, Denver, CO  
[06/2016, 12/2015] The CTO Office of EMC CTD, Princeton, NJ
- **Pricing Games for Hybrid Object Stores in the Cloud: Provider vs. Tenant**  
[07/2015] USENIX HotCloud'15, Santa Clara, CA  
[06/2015] The CTO Office of EMC CTD, Princeton, NJ
- **CAST: Tiering Storage for Data Analytics in the Cloud**  
[06/2015] ACM HPDC'15, Portland, OR
- **An In-Memory Object Caching Framework with Adaptive Load Balancing**  
[04/2015] ACM EuroSys'15, Bordeaux, France  
[08/2014] IBM Almaden Research Center, San Jose, CA
- **High Performance, Flexible Memory Caching**  
[08/2013] IBM Almaden Research Center, San Jose, CA

---

## Professional Services

- **Program Committee:** BDCAT'17
- **Shadow PC:** EuroSys'16/17
- **External Reviewer:** BigData'14/16, HPDC'14/15/16/17, ICDCS'16/17, ICPP'16, SC'15/16, ICS'17, NAS'17, Cluster'17
- **Session Summarizer:** FAST'14