

Mehmet Fatih Aktaş

+1 (732) 640 8711 ✉ mfatihaktas@gmail.com 🌐 [mfatihaktas](#) 📄 [mfatihaktas](#)

Research Interests

- Distributed Systems Probabilistic Modeling Performance Optimization Reinforcement Learning
- Worked on both theoretical analysis and development (see [github](#)) of large scale storage and computing systems.
 - In particular on distributed in-memory data storage, straggler mitigation in distributed computing, load balancing in distributed storage, and resource provisioning in the Cloud.
 - Most recently on scheduling user requests across multiple edge computing clusters to minimize response time.

Education

PhD in Electrical & Computer Engineering 2015-2020
Rutgers University, Piscataway, NJ GPA: 4.0/4.0
Dissertation: Performance evaluation of redundancy techniques for distributed storage and computing systems
Advisor: Prof. Emina Soljanin

Masters in Electrical & Computer Engineering 2012-2015
Rutgers University, Piscataway, NJ GPA: 4.0/4.0
Thesis: Scheduling and flexible control of wide-area data transport services for end-to-end application workflows
Advisor: Prof. Manish Parashar

Bachelors in Electrical & Electronics Engineering 2007-2012
Bilkent University, Ankara, Turkey GPA: 3.84/4.0

Experience

Assistant Professor – Computer Science, Bilkent University Since August 2021
– The goal of my research is to make distributed computing systems faster and more robust to runtime changes.

Postdoctoral Associate – WINLAB, Rutgers University March – August 2021
Scheduling (e.g., image processing requests) across edge clusters for low-latency processing
– Developed a distributed scheduling algorithm to assign user requests across the edge clusters. Built a [prototype](#) and demonstrated that our policy is robust against fluctuating load and runtime changes in the system [P1].

Senior Software Engineer – MathWorks, Natick, MA March 2020 – March 2021
Member of language execution team
– Worked on the dynamic compilation and execution of Matlab. Acted as the lead designer and developer of the first code coverage tool for Matlab.

Research Intern – IBM Research, Yorktown Heights, NY Summer 2018
An autonomic container sizing system for Kubernetes. (jointly with Chen Wang, Alaa Youssef)
– Proposed a system to predict and refine the CPU and Memory reservations of containers in [Kubernetes](#).
– Showed, with a cluster trace driven analysis and simulation, that our proposed system can raise the average CPU (memory) utilization in the cluster from 10% (26%) to 54% (88%) [C6, P4].

Research Intern – Argonne National Laboratory, Lemont, IL Summer 2016
Multi-layered intermediate buffering for in-situ workflows. (jointly with Tom Peterka)
– Developed a multi-layer data staging service within [Decaf](#); a distributed workflow execution framework.
– Integrated Decaf with [DataSpaces](#); a distributed in-memory data staging framework for HPC.

Research Assistant – Rutgers University, Piscataway, NJ Fall 2012 – Spring 2020
Funded by the university or NSF to do research during my Masters and PhD studies.

Key Research Projects

Straggler mitigation with redundancy in distributed computing Spring 2017-present
What is the gain and pain of using redundancy for straggler mitigation?
– Analyzed the cost (pain) and latency (gain) of introducing early or delayed redundancy (replicated or coded) in execution of jobs with multiple parallel tasks (e.g., MapReduce) [C9].
– Extended the analysis for straggler relaunch and validated results using Google cluster data [C10].
– Developed a model that closely captures the effect of added redundancy on system response time and extended

the previous analysis for this more realistic setting [J3].

Scheduling redundancy for Master-Slave compute clusters

Spring 2018-present

What is the right strategy to schedule jobs of many tasks with redundancy?

- With Deep Q-learning, we learn the characteristics of “good” policies for scheduling jobs with redundancy.
- With Queueing Theory, we propose and analyze a policy that mimics the learned principles, and show that our model driven policy performs as good as the more complex policies learned by Deep-RL [C3, P2]

Redundancy for robust service in distributed storage

Fall 2017-present

Can we build a distributed storage that is robust to skews in content popularity?

- Proved the robustness of binary simplex codes against the skews in content popularity [C11].
- Showed that in a system of n storage nodes, when each object is replicated in d nodes, system’s load balancing ability increases multiplicatively with d as long as $d = o(\log(n))$, increases exponentially as soon as $d = \Theta(\log(n))$.
- Showed that using r -XOR’s to implement d choices for each object rather than replication reduces storage overhead multiplicatively with r , while also reducing system’s load balancing ability additively with r [C1, J1].

Content download from coded distributed storage

Fall 2016-present

Can we use the storage redundancy for fast content access?

- Analyzed the performance of download with redundant requests in storage systems which employ availability codes [C8, J2].
- Analyzed download time in MDS coded storage with minimal parallelism [C5].

Anonymity vs. Delay tradeoff in Mix networks

Summer 2017-present

What is the delay penalty of anonymizing senders at the network switches?

- Extended an approximate method for analyzing assembly-like queues to study the delay in batch mixes.
- Proposed a randomized batch scheme that improves the delay exponentially at the cost of giving up on the strict anonymity guarantees of the deterministic batch mix [C2].
- Currently trying to answer the question: How long does it take to deanonymize ToR users with intersection attacks? [P3].

Wide-area data staging and sharing

Spring 2015-Spring 2016

Can we build a wide-area scale big memory abstraction?

- Developed [WA-DataSpaces](#); a big shared memory abstraction for applications distributed across the globe.
- Our framework was used in a [showcase](#) entry titled “InfiniCortex” at SC’14 and SC’15.
- Demonstrated the effectiveness of probabilistic prefetchers for on-time data locality in scientific applications [C7].

Opportunistic in-transit data processing and staging

Fall 2012-Fall 2014

Can SDN technology enable just-in-time end-to-end data delivery?

- Designed and developed a Network OS prototype for intermediate data staging and processing [C12].
- Demonstrated that scheduling of intermediate resources with a disciplined convex optimization program can enable just-in-time end-to-end data delivery [J4].

Graduate Courses Taken

Stochastic Processes	Queueing Theory	Computer Networks
Error Control Coding	Science of System Resilience	Linear System Analysis
Convex Optimization	Detection and Estimation Theory	Machine Learning
Parallel and Distributed Computing	Advanced Computer Architecture	Software Engineering

Courses Taught

Probability and Random Processes (as Instructor, TA)	Computer Networks (as Professor)
Programming Methodology	Introduction to Algorithms and Programming
Parallel and Distributed Programming (Graduate level)	Computer Architecture

Software Skills

Programming: Fluent in Python and C++.

Algorithms: Numpy, Scipy, Statsmodels, Mpmath, NetworkX, Matplotlib, TensorFlow, C++/Boost.

Systems: Docker containers, Kubernetes, Terraform, Envoy, Flask, MPI, Simpy (for system simulations).

Parallelism: OpenMP, POSIX threads, Python3-Concurrency (threading, asyncio), Boost (thread, asio).

Networking: Python3 (socket, socketserver), RDMA IB Verbs, POX, FlowVisor, Open vSwitch, Mininet.

Visualization: Grafana, Kibana, Elastic stack (used them to search and monitor container resource usage metrics).

Predictive analytics: Knowledgeable in Markov models, time-series prediction (e.g. ARIMA) and neural networks.

Awards

Academic achievement award from Rutgers University

Full scholarship from Bilkent University

Journal Papers

Published.....

- [J1] Mehmet F. Aktaş, Amir Behrouzi Far, Emina Soljanin, and Philip Whiting. "Evaluating Load Balancing Performance in Distributed Storage with Redundancy". In: *IEEE Transactions on Information Theory* (2021). DOI: [10.1109/TIT.2021.3054385](https://doi.org/10.1109/TIT.2021.3054385). [arXiv:1910.05791](https://arxiv.org/abs/1910.05791).
- [J2] Mehmet F. Aktaş, Swanand Kadhe, Emina Soljanin, and Alex Sprintson. "Download Time Analysis for Distributed Storage Codes with Locality and Availability". In: *IEEE Transactions on Communications* (Sept. 2020). [arXiv:1912.09765](https://arxiv.org/abs/1912.09765).
- [J3] Mehmet F. Aktaş and Emina Soljanin. "Straggler Mitigation at Scale". In: *IEEE/ACM Transactions on Networking* 27 (Oct. 2019), pp. 2266–2279. [arXiv:1906.10664](https://arxiv.org/abs/1906.10664).
- [J4] Mehmet F. Aktaş, Georgiana Haldeman, and Manish Parashar. "Scheduling and flexible control of bandwidth and in-transit services for end-to-end application workflows". In: *Future Generation Computer Systems* 56 (Mar. 2016), pp. 284–294. Elsevier.

Accepted.....

- [S1] Mehmet F. Aktaş, Gauri Joshi, Swanand Kadhe, Fatemeh Kazemi, and Emina Soljanin. "Service Rate Region: A New Aspect of Coded Distributed System Design". In: *IEEE Transactions on Information Theory* (Sept. 2021). [arXiv:2009.01598](https://arxiv.org/abs/2009.01598).

In preparation.....

- [P1] Mehmet F. Aktaş and Dipankar Raychaudhuri. "Distributed client-side scheduling for edge computing". In: ... (2021).
- [P2] Mehmet F. Aktaş and Emina Soljanin. "Optimizing Redundancy Levels in Master-Worker Compute Clusters for Straggler Mitigation". In: ... (2021). [arXiv:1906.05345](https://arxiv.org/abs/1906.05345).
- [P3] Mehmet F. Aktaş, Mansi Sood, Allison Beemer, and Emina Soljanin. "How long does it take to deanonymize ToR users with intersection attack?" In: ... (2021).
- [P4] Mehmet F. Aktaş, Chen Wang, Alaa Youssef, and Malgorzata (Gosia) Steinder. "kAdvisor: An autonomic container sizing system for Kubernetes". In: *IEEE Transactions on Cloud Computing* (2021).

Refereed Conference Papers

- [C1] Mehmet F. Aktaş, Amir Behrouzi-Far, and Emina Soljanin. "Load Balancing Performance in Distributed Storage with Regular Balanced Redundancy". In: *Redundancy 2019: International Symposium on Problems of Redundancy in Information and Control Systems*. IEEE. 2019.
- [C2] Mehmet F. Aktaş and Emina Soljanin. "Anonymity Mixes as (Partial) Assembly Queues: Modeling and Analysis". In: *Information Theory Workshop*. IEEE. 2019. [arXiv:1907.11603](https://arxiv.org/abs/1907.11603).
- [C3] Mehmet F. Aktaş and Emina Soljanin. "Learning Effective Straggler Mitigation from Experience and Modeling". In: *CodML Workshop of ICML*. ACM. 2019.
- [C4] Gala Yadgar, Oleg Kolosov, Mehmet F. Aktaş, and Emina Soljanin. "Modeling The Edge: Peer-to-Peer Reincarnated". In: *USENIX Workshop on Hot Topics in Edge Computing*. USENIX. 2019.
- [C5] Mehmet F. Aktaş and Emina Soljanin. "Heuristics for Analyzing Download Time in MDS Coded Storage Systems". In: *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2018.
- [C6] Mehmet F. Aktaş, Chen Wang, Alaa Youssef, and Malgorzata (Gosia) Steinder. "Resource Profile Advisor for Containers in Cognitive Platform". In: *Symposium on Cloud Computing*. ACM. 2018.

- [C7] Mehmet F. Aktaş, Javier Diaz-Montes, Ivan Rodero, and Manish Parashar. "WA-Dataspace: Exploring the Data Staging Abstractions for Wide-Area Distributed Scientific Workflows". In: *Proceedings of the 46th International Conference on Parallel Processing*. ACM. 2017.
- [C8] Mehmet F. Aktaş, Elie Najm, and Emina Soljanin. "Simplex Queues for Hot-Data Download". In: *Proceedings of the SIGMETRICS/International Conference on Measurement and Modeling of Computer Systems*. ACM. 2017.
- [C9] Mehmet F. Aktaş, Pei Peng, and Emina Soljanin. "Effective Straggler Mitigation: Which Clones Should Attack and When?" In: *MAMA Workshop of Sigmetrics*. ACM. 2017. [arXiv:1710.00748](https://arxiv.org/abs/1710.00748).
- [C10] Mehmet F. Aktaş, Pei Peng, and Emina Soljanin. "Straggler Mitigation by Delayed Relaunch of Tasks". In: *Proceedings of the IFIP WG 7.3 Performance*. ACM. 2017. [arXiv:1710.00414](https://arxiv.org/abs/1710.00414).
- [C11] Mehmet F. Aktaş et al. "On the Service Capacity Region of Accessing Erasure Coded Content". In: *55th Annual Allerton Conference on Communication, Control, and Computing*. 2017. [arXiv:1710.03376](https://arxiv.org/abs/1710.03376).
- [C12] Mehmet F. Aktaş, Georgiana Haldeman, and Manish Parashar. "Flexible Scheduling and Control of Bandwidth and in-transit Services for end-to-end Application Workflows". In: *Proceedings of the Fourth International Workshop on Network-Aware Data Management of Supercomputing Conference*. IEEE Press. 2014.