

George Papadimitriou

RESEARCH INTERESTS

My interests lie within the intersection of Distributed Computing and Data Intensive Applications. This includes concepts manifesting in High Performance Computing, Cloud Computing and Big Data systems.

EDUCATION

AUG. 2017	UNIVERSITY OF SOUTHERN CALIFORNIA (USC), USA
⋮	<i>Viterbi School of Engineering</i>
	<i>Degree Pursued:</i> PhD in Computer Science
NOW	<i>Advisor:</i> Ewa Deelman
	<i>Current Applied GPA:</i> 3.76/4.00
	<i>Current Total GPA:</i> 3.62/4.00
SEP. 2010	NATIONAL TECHNICAL UNIVERSITY OF ATHENS (NTUA), Greece
⋮	<i>Electrical and Computer Engineering (ECE)</i>
	<i>Degree:</i> Diploma in Electrical and Computer Engineering (Integrated Master)
FEB. 2018	<i>GPA:</i> 8.06/10.0

AWARDS

AUG. 2021	<i>Editor's Choice Article.</i> My article published in FGCS journal titled "End-to-end online performance data capture and analysis for scientific workflows", was selected as an editor's choice article.
JUL. 2020	<i>PEARC'20 Best Student Paper.</i> Best student paper in "Advanced research computing environments – systems and system software" Track, for the publication "Workflow Submit Nodes as a Service on Leadership Class Systems"
NOV. 2019	<i>SCinet Tech Challenge 2019 - Most Diverse Resource Set.</i> Received the award for the most diverse resource set used during the SCinet Tech Challenge 2019 at the Supercomputing Conference '19.
NOV. 2019	<i>SCinet Tech Challenge 2019 - Most Original Technical Approach.</i> Received the award for the most original technical approach used during the SCinet Tech Challenge 2019 at the Supercomputing Conference '19.
OCT. 2019	<i>Gerondelis Foundation Scholarship.</i> The Gerondelis Foundation, Inc. awarded me the Gerondelis scholarship of \$5,000 for the 2019-2020 academic year.
AUG. 2017	<i>ISI Distinguished Top-Off Fellowship.</i> The Information Sciences Institute, awarded me the total amount of \$5,000 under the ISI Top-Off Fellowship for the 2017-2018 academic year.

WORK EXPERIENCE

AUG. 2017	INFORMATION SCIENCES INSTITUTE – <i>Graduate Research Assistant</i>
⋮	I'm part of the SciTech group, working on the projects: Panorama360 , Dynamo and FlyNet.
	I'm currently working on the characterization of machine learning science workflows.
NOW	Tools: Pegasus WMS, Python, InfluxDB, RabbitMQ, Shell Scripting, Git
JUNE 2018	OAK RIDGE NATIONAL LABORATORY – <i>ASTRO Internship</i>
⋮	I was part of the Future Technologies Group, collaborating with Spallation Neutron Source scientists and building new workflows.
SEPT. 2018	Tools: Pegasus WMS, Docker, Singularity, Python, Shell Scripting, Git
OCT. 2014	CENTENTIA S.A – <i>Full Stack Junior Software Engineer</i>
⋮	I worked with the AroTRON team, designing and developing components of Cententia's eCRM solution.
	Core developer of the MIS Report Designer and Data Integration Modules.
OCT. 2016	Tools: C#, SQL Server, HTML, CSS, Javascript, Kendo UI, Perforce, Visual Studio, JIRA, Confluence

TEACHING EXPERIENCE

- | | |
|-----------|--|
| FALL 2021 | Teaching assistant for the undergraduate course <i>CSC350 Introduction to Operating Systems</i> , Viterbi School of Engineering, USC <ul style="list-style-type: none">- Assisting students with their programming assignments- Creating exam questions and grading exams |
| FALL 2015 | Lab assistant for the undergraduate course of <i>Introduction to Programming</i> , school of ECE, NTUA <ul style="list-style-type: none">- Assisting students with their programming assignments. |
| FALL 2014 | Lab assistant for the undergraduate course of <i>Operating Systems</i> , school of ECE, NTUA <ul style="list-style-type: none">- Assisting students with course material and their programming assignments. |

PROJECTS

- **Panorama360**

While scientific workflows have become extremely complex, existing tools only allow for capturing performance statistics of some parts of them in a non-unified form. Thus, synthesizing a complete picture of how a workflow performs is almost impossible. Panorama360, by using the Pegasus WMS aims to provide a way to the community to automatically collect and store end-to-end performance traces of their workflows in a common JSON format. Additionally, among Panorama360's goals is to offer new ways of interpreting these data, using Machine Learning Techniques. My role in this project is to create the data collection architecture and the open access data repository.

- **Dynamo**

DyNamo is a project funded under the NSF Campus Cyberinfrastructure program, which aims to enable high-performance, adaptive, and performance-isolated data-flows across a federation of distributed cloud resources and community data repositories. Using the Mobius platform, DyNamo facilitates the provisioning of appropriate compute and storage resources for observational science workflows from diverse, national-scale cyberinfrastructure (CI). In this project I'm working to enable the automatic orchestration of data-driven workflows on dynamically provisioned resources using the Pegasus Workflow Management System.

- **NTUA Diploma Thesis**

The goal of my Diploma thesis was the design of an anomaly detection system for compute nodes. In the course of this project, I evaluated the application of existing distributed implementations of anomaly detection algorithms bundled in the Apache's Spark [MLlib](#). Then I tested and implemented a version of the Robust PCA algorithm in Spark, that was staging in data from an HBASE cluster and annotating anomalies on its output.

Tools: [Apache Spark](#), [Apache HBase](#)

COURSE PROJECTS

- | | |
|-------------|--|
| FALL 2018 | “Advanced Operating Systems” (USC)
Built a simple operating systems named JOS that is based on an exokernel approach.
Technologies Used: C, QEMU-KVM, Git |
| FALL 2017 | “Advanced Data Stores” (USC)
Evaluation of Oracle's Sharding and JSON capabilities with the use of the YCSB and NoBench benchmarks.
Technologies Used: Python, Java, Git |
| SPRING 2015 | “Software Engineering” (NTUA)
Addition of new functionalities to the SIP Communicator and JAIN SIP Proxy server.
Technologies Used: Java, MySQL, Git |
| FALL 2014 | “Programming Languages II” (NTUA)
Implementation of a type inference system for the simply typed lambda calculus.
Technologies Used: Haskell, Linux, Shell Scripting |
| FALL 2014 | “Parallel Processing Systems” (NTUA)
Implementation and performance analysis of parallel algorithms that solve Thermal Equilibrium.
Technologies Used: C, Linux, MPI, OpenMP, CUDA, Shell Scripting, Git |

SPRING 2014	<p>“Compilers” (NTUA)</p> <p>Implementation of lexical and syntactical analysis for the imperative C-like programming language Pazcal^b.</p> <p>Technologies Used: OCaml, ocamllex, ocaml yacc, Shell Scripting, Git</p>
SPRING 2014	<p>“Operating Systems Laboratory” (NTUA)</p> <p>Linux device driver modification and implementation of a virtual crypto device in QEMU-KVM.</p> <p>Technologies Used: C, Linux, QEMU-KVM, VirtIO, Shell Scripting, Git</p>

SKILLS

Programming Languages:	C/C++, C#, Java, OCaml, Haskell, Scala, Python, PHP, Bash Scripting, SQL
Web Technologies:	HTML, CSS, Javascript, Ajax, Kendo UI, Bootstrap
Databases Systems:	SQL (MySQL, SQL Server), NoSQL (HBase, MongoDB, InfluxDB)
Operating Systems:	GNU/Linux, MS Windows
Tools:	Git, GDB, Valgrind, VIM, Visual Studio, L ^A T _E X, JIRA, Confluence, Wireshark

LANGUAGES

GREEK:	Mother tongue
ENGLISH:	IELTS (June 2016): Band 7.5 University of Michigan ECPE — CEFR Level C2
FRENCH:	Diplôme d'études en langue française (DEL F) — CEFR Level B1

PUBLICATIONS

-
- [1] Patrycja Krawczuk, George Papadimitriou, Shubham Nagarkar, Mariam Kiran, Anirban Mandal, and Ewa Deelman. “Anomaly Detection in Scientific Workflows using End-to-End Execution Gantt Charts and Convolutional Neural Networks”. In: *Proceedings of the Practice and Experience in Advanced Research Computing*. PEARC '21. Funding Acknowledgments: DOE DESC0012636. Boston, MA, USA: Association for Computing Machinery, 2021. ISBN: 978-1-4503-8292-2/21/07. DOI: [10.1145/3437359.3465597](https://doi.org/10.1145/3437359.3465597). URL: <http://doi.acm.org/10.1145/3437359.3465597>.
 - [2] George Papadimitriou, Cong Wang, Karan Vahi, Rafael Ferreira da Silva, Anirban Mandal, Liu Zhengchun, Rajiv Mayani, Mats Rynge, Mariam Kiran, Vickie E. Lynch, Rajkumar Kettimuthu, Ewa Deelman, Jeffrey S. Vetter, and Ian Foster. “End-to-End Online Performance Data Capture and Analysis for Scientific Workflows”. In: *Future Generation Computer Systems* 117 (2021). Funding Acknowledgments: DOE DE-SC0012636, pp. 387–400. ISSN: 0167-739X. DOI: <https://doi.org/10.1016/j.future.2020.11.024>. URL: <http://www.sciencedirect.com/science/article/pii/S0167739X20330570>.
 - [3] H. Tu, G. Papadimitriou, M. Kiran, C. Wang, A. Mandal, E. Deelman, and T. Menzies. “Mining Workflows for Anomalous Data Transfers”. In: *2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR) (MSR)*. Funding Acknowledgment: NSF 1826574, NSF 1931425, DOE DE-SC0012636. Los Alamitos, CA, USA: IEEE Computer Society, May 2021, pp. 1–12. DOI: [10.1109/MSR52588.2021.00013](https://doi.ieeecomputersociety.org/10.1109/MSR52588.2021.00013). URL: <https://doi.ieeecomputersociety.org/10.1109/MSR52588.2021.00013>.
 - [4] Mariam Kiran, Cong Wang, George Papadimitriou, Anirban Mandal, and Ewa Deelman. “Detecting Anomalous Packets in Network Transfers: Investigations using PCA, Autoencoder and Isolation Forest in TCP”. In: *Machine Learning* (2020). Funding Acknowledgments: DOE DESC0012636. ISSN: 1573-0565. DOI: [10.1007/s10994-020-05870-y](https://doi.org/10.1007/s10994-020-05870-y). URL: <https://doi.org/10.1007/s10994-020-05870-y>.
 - [5] Eric Lyons, David Westbrook, Andrew Grote, George Papadimitriou, Komal Thareja, Cong Wang, Michael Zink, Ewa Deelman, Anirban Mandal, and Paul Ruth. “An On-Demand Weather Avoidance System for Small Aircraft Flight Path Routing”. In: *Dynamic Data Driven Application Systems*. Ed. by Frederica Darema, Erik Blasch, Sai Ravela, and Alex Aved. Funding Acknowledgments: NSF 1826997, 2018074. Cham: Springer International Publishing, 2020, pp. 311–319. ISBN: 978-3-030-61725-7. DOI: [10.1007/978-3-030-61725-7_36](https://doi.org/10.1007/978-3-030-61725-7_36).
 - [6] Eric Lyons, Michael Zink, Anirban Mandal, Cong Wang, Paul Ruth, Chandrasekar Radhakrishnan, George Papadimitriou, Ewa Deelman, Komal Thareja, and Ivan Rodero. “DyName: Scalable Weather Workflow Processing in the Academic MultiCloud”. In: *100th American Meteorological Society Annual Meeting* (2020). Funding Acknowledgments: NSF 1826997.
 - [7] George Papadimitriou, Eric Lyons, Cong Wang, Komal Thareja, Ryan Tanaka, Paul Ruth, J.J. Villalobos, Ivan Rodero, Ewa Deelman, Michael Zink, and Anirban Mandal. “Application Aware Software Defined Flows of Workflow Ensembles”. In: *2020 IEEE/ACM Innovating the Network for Data-Intensive Science (INDIS)*. Funding Acknowledgments: NSF 1826997. 2020, pp. 10–21. DOI: [10.1109/INDIS51933.2020.00007](https://doi.org/10.1109/INDIS51933.2020.00007).
 - [8] George Papadimitriou, Karan Vahi, Jason Kincl, Valentine Anantharaj, Ewa Deelman, and Jack Wells. “Workflow Submit Nodes as a Service on Leadership Class Systems”. In: *Proceedings of the Practice and Experience*

- in *Advanced Research Computing*. PEARC '20. Funding Acknowledgments: DOE DESC0012636. Portland, OR, USA: Association for Computing Machinery, 2020. ISBN: 978-1-4503-6689-2/20/07. DOI: [10.1145/3311790.3396671](https://doi.org/10.1145/3311790.3396671). URL: <http://doi.acm.org/10.1145/3311790.3396671>.
- [9] Karan Vahi, Danny Goldstein, George Papadimitriou, Peter Nugent, and Ewa Deelman. “Gearing the DECam Analysis Pipeline for Multi-Messenger Astronomy using Pegasus Workflows”. In: *Astronomical Data Analysis Software and Systems (ADASS) XXIX* (2020). Funding Acknowledgments: NSF 1664162, DOE DESC0012636.
 - [10] Cong Wang, George Papadimitriou, Mariam Kiran, Anirban Mandal, and Ewa Deelman. “Identifying Execution Anomalies for Data Intensive Workflows Using Lightweight ML Techniques”. In: *2020 IEEE High Performance Extreme Computing Conference (HPEC)*. Funding Acknowledgments: DOE DESC0012636. 2020, pp. 1–7. DOI: [10.1109/HPEC43674.2020.9286139](https://doi.org/10.1109/HPEC43674.2020.9286139).
 - [11] Ewa Deelman, Karan Vahi, Mats Rynge, Rajiv Mayani, Rafael Ferreira da Silva, George Papadimitriou, and Miron Livny. “The Evolution of the Pegasus Workflow Management Software”. In: *Computing in Science Engineering* 21.4 (2019). Funding Acknowledgments: NSF 1664162, NSF 1148515, DOE DESC0012636, NSF 1642053, pp. 22–36. DOI: [10.1109/MCSE.2019.2919690](https://doi.org/10.1109/MCSE.2019.2919690).
 - [12] Rafael Ferreira da Silva, Scott Callaghan, Tu Mai Anh Do, George Papadimitriou, and Ewa Deelman. “Measuring the Impact of Burst Buffers on Data-Intensive Scientific Workflows”. In: *Future Generation Computer Systems* 101 (2019). Funding Acknowledgments: DOE DESC0012636, NSF 1664162, NSF 1741040, pp. 208–220. DOI: [10.1016/j.future.2019.06.016](https://doi.org/10.1016/j.future.2019.06.016).
 - [13] Eric Lyons, George Papadimitriou, Cong Wang, Komal Thareja, Paul Ruth, J.J. Villalobos, Ivan Rodero, Ewa Deelman, Michael Zink, and Anirban Mandal. “Toward a Dynamic Network-centric Distributed Cloud Platform for Scientific Workflows: A Case Study for Adaptive Weather Sensing”. In: *15th International Conference on eScience (eScience)*. Funding Acknowledgments: NSF 1826997. San Diego, CA, USA, 2019, pp. 67–76. DOI: [10.1109/eScience.2019.00015](https://doi.org/10.1109/eScience.2019.00015).
 - [14] George Papadimitriou, Mariam Kiran, Cong Wang, Anirban Mandal, and Ewa Deelman. “Training Classifiers to Identify TCP Signatures in Scientific Workflows”. In: *2019 IEEE/ACM Innovating the Network for Data-Intensive Science (INDIS)*. Funding Acknowledgments: DOE DESC0012636. Denver, CO, USA, 2019, pp. 61–68. DOI: [10.1109/INDIS49552.2019.00012](https://doi.org/10.1109/INDIS49552.2019.00012).
 - [15] Ivan Rodero, Yubo Qin, Jesus Valls, Anthony Simonet, J.J. Villalobos, Manish Parashar, Chooban Youn, Cong Wang, Komal Thareja, Paul Ruth, George Papadimitriou, Eric Lyons, and Michael Zink. “Enabling Data Streaming-based Science Gateway through Federated Cyberinfrastructure”. In: *Gateways 2019*. Funding Acknowledgments: NSF 1835692, NSF 1745246, NSF 1826997. San Diego, CA, USA, 2019.
 - [16] Karan Vahi, Mats Rynge, George Papadimitriou, Duncan Brown, Rajiv Mayani, Rafael Ferreira da Silva, Ewa Deelman, Anirban Mandal, Eric Lyons, and Michael Zink. “Custom Execution Environments with Containers in Pegasus-enabled Scientific Workflows”. In: *15th International Conference on eScience (eScience)*. Funding Acknowledgments: NSF 1664162, NSF 1826997, NSF 1443047. San Diego, CA, USA, 2019, pp. 281–290. DOI: [10.1109/eScience.2019.00039](https://doi.org/10.1109/eScience.2019.00039).