Curriculum Vitae

Amir H. Payberah

1 Personal Data

1.1 Name

- **Name:** Amir H. Payberah
- **Date of birth:** 27 August, 1978
- **Gender:** Male

1.2 Contact Address

- **Phone:** +46-(0)72-55 44 011
- **Email:** payberah@kth.se
- **Web:** https://payberah.github.io
- **Workplace address:** Isafjordsgatan 22, 16440 Kista, Sweden

1.3 Present Employment

- July 2018 - Now, Assistant Professor in Computer Science (Software Systems), School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Kista, Sweden.

1.4 Previous Employment

- April 2017 - June 2018, Machine Learning Scientist, University of Oxford, Oxford, UK.
  Developing machine learning and deep learning algorithms to find patterns and trends in large biomedical datasets.
- July 2013 - May 2017, Senior Researcher, SICS Swedish ICT, Kista, Sweden.
  Worked on mining massive datasets and data intensive computing platforms (mainly Spark), and on data streaming and graph processing algorithms.
- August 2016 - December 2016, Adjunct Lecturer, KTH Royal Institute of Technology, Kista, Sweden (part time, 20%).
  Designed and gave a course on Data Intensive Computing Platforms for a new Masters programme in data science. The course covered a wide range of topics, including (i) different distributed filesystems and NoSQL databases, (ii) different processing models for various data

types, e.g., batch-data, streaming-data, graph-data, and structured-data, and (iii) different resource management systems.

- August 2008 - June 2013, PhD Student, KTH Royal Institute of Technology, Kista, Sweden.
  Worked on distributed systems, including Peer-to-Peer content distribution networks, gossip-based algorithms and large scale publish/subscribe systems.

- January 2007 - April 2007, Researcher, Parallel Dator Centrum (PDC), Stockholm, Sweden.
  Worked on grid resource management, and dynamic resource allocation on gLite.

- January 2000 - December 2006, Linux Technical Leader, DPI Co., SiNA Co., and GATA Co., Tehran, Iran.
  Built different customized Linux distributions, e.g., embedded Linux kernel, file systems, and device drivers.

- January 2001 - December 2002, Network Administrator, DP Co., Tehran, Iran
  Worked as a local area network administrator.

2 Education Qualification

- 2008 - 2013, PhD in Information and Communication Technology, KTH Royal Institute of Technology, Kista, Sweden.
  Thesis: Live streaming in P2P and hybrid P2P-Cloud environments for the open Internet.

- 2006 - 2008, MSc in Software Engineering of Distributed Systems, KTH Royal Institute of Technology, Kista, Sweden.
  Thesis: ForestCast for P2P live streaming.

- 2000 - 2003, MSc in Computer Architecture, Amirkabir University of Technology (AUT), Tehran, Iran.
  Thesis: Distributed inspection of intrusion detection systems using mobile agents.

- 1996 - 2000, BSc in Computer Engineering, Amirkabir University of Technology (AUT), Tehran, Iran.
  Thesis: Designing and implementing an I/O card and developing its device driver in GNU/Linux.

3 Scientific Achievements

My research interests cover a variety of topics in distributed algorithms, data intensive computing, and data mining. During my Postdoc studies, I conducted research on mining massive datasets and data intensive computing platforms, and in my PhD studies, I worked on distribute algorithms, and Peer-to-Peer overlay networks. In this section, I will elaborate on my research profile. References of the form [.] refer to the publications, listed in the publication section.

Data Intensive Computing (2011 - Now)
My works on this topic are divided into two main parts:

- **Big Data Analytics (2013 - Now)** Since September 2017, I have been working on a project called Deep Medicine. Deep Medicine is a new initiative of the Oxford Martin School at the

University of Oxford with the goal of developing a Big Data analytics platform to solve high-impact health care problems, with the application of modern machine learning algorithms to large multi-modal (e.g., genetics, imaging, medical records) biomedical datasets (including, but not limited to, UK BioBank). I use existing machine learning and deep learning algorithms, and develop new ones, to find patterns and trends in large datasets.

Prior to that, in a project, called BIDAF (Big Data Analysis Framework) at SICS, I worked on applying machine learning techniques on massive data, and I conducted research on evolutionary clustering of large dataset. In evolutionary clustering, we aim to cluster continuously changing data, while new clusters should not deviate dramatically from the recent history. As a part of this project, we also enriched some of the existing mining algorithms to enable them to work with big data. One of these algorithms is Growing Neural Gas (GNG) to handle concept drifting of data streams. This technique can be used for different purposes, such as novelty detection or clustering. We made a parallel model of the algorithm in [1], such that it can be implemented in any data-parallel processing platform, e.g., Spark.

I also worked within another project called DOIT (Data-Driven Optimization for Intelligent and Efficient Transports), a joint work with Scania. The goal of this project was to demonstrate the effectiveness of big data analytics and optimization in improving the decision support and reducing the cost in heavy-duty road transportations. In this project, we applied different supervised learning methods, e.g., regressions, decision tree, and random forest, to build models for predicting fuel consumption based on the data of existing trips. All the implementations were based on MLlib, the Spark machine learning library.

Similarly, in another project, called TrafMod (Traffic Modeling), we worked with Ericsson to make use of big data analytics, data management and machine learning and set up a traffic modeling pipeline to quickly and automatically turn massive amounts of data collected from the network into functional traffic models. We built a platform using Spark and Lift (a web framework for Scala), and since Spark machine learning libraries were not available back then (early 2013), we ourselves implemented different data clustering algorithms, such as KMeans clustering, in Spark.

- **Distributed Partitioning of Massive Graphs (2011 - 2017).** Everyday, tremendous amount of data is generated through online services, social networks, and wearable devices. Most of this data can be described as graphs and can be exploited with a rich set of algorithms, e.g., community detection, trend detection and anomaly detection. There exist a number of distributed graph processing systems (e.g., GraphLab, GraphX, and Giraph) for developing such algorithms. However, the performance of executing graph algorithms on such systems is tightly coupled with the way the input graph is partitioned across multiple machines in a cluster.

  Graph partitioning is a well known NP-complete problem with a wide range of applications and solutions. In very large-scale distributed scenarios, state-of-the-art partitioning algorithms are not directly applicable, because they typically involve frequent global operations over the entire graph. To overcome these challenges, we proposed a fully distributed algorithm, called Ja-Be-Ja [2, 6, 7], that uses local search and simulated annealing techniques for edge-cut partitioning. The algorithm is massively parallel: there is no central coordination, each node is processed independently, and only the direct neighbors of the node, and a small subset of random nodes in the graph need to be known locally. The paper [7] got the **best paper award** from the IEEE international conference on Self-Adaptive and Self-Organizing Systems (SASO) 2013. This work was later extended to include both edge-cut and vertex-cut partitioning [2].

I continued my research in this area by taking into account the streaming nature of many large graphs generated through online services. In this data model, the graph elements (vertices or edges) arrive sequentially over time and the computation should be done based on the received element properties on-the-fly. A few data streaming algorithms have been recently developed for centralized data stream processing, where data stream is sent from a single source. Nevertheless, collecting data from a single source is not always possible for variety of reasons, e.g., bottlenecks or privacy issues. To overcome the challenges imposed by this constraint, we designed and implemented HoverCut [4], a distributed algorithm for vertex-cut partitioning of streaming graphs that uses multi-threading/multi-processing together with a windowing technique to bring about a light-weight state sharing between threads/processes. This work was published in IEEE BigData Congress 2016 and received the **best paper award**.

Distributed Systems (2008 - 2015)
Peer-to-Peer (P2P) systems are widely used over the Internet. In the last decade, a lot of research efforts have been made in this area, and now we are witnessing their achievements in real systems. Video streaming and file sharing are two successful applications of the P2P technology. My work on P2P systems includes:

- **P2P Live Media Streaming (2008 - 2014).** In P2P live streaming systems, nodes should receive the stream with respect to timing constraints, while the P2P overlay adapts to the changes in the network. Moreover, nodes should be incentivized to contribute their resources. I addressed these problems in [3, 13, 17] by introducing a distributed auction algorithm to build efficient overlays for live media streaming in form of a tree [17, 18] and a mesh [13]. These algorithms are fully distributed, such that each node has only a partial information about the system that it acquires by participating in a gossiping algorithm.

  Moreover, due to the high volume of video content, P2P overlays often face bandwidth bottleneck, which results in playback latency and playback continuity. I addressed this problem by proposing a hybrid P2P-Cloud model, called CLive [10], to satisfy real-time constraints on delay between the generation of the stream and its actual delivery to users, and to maintain the QoS with the least possible cost. The idea is to borrow some resources from the Cloud, upon need, to increase the amount of total available bandwidth. However, as the use of Cloud resources costs money, I modeled the problem as a minimization of the economical cost, provided that a set of constraints on QoS is satisfied.

- **Distributed NAT Traversal (2010 - 2015).** Existing Network Address Translation gateways (NATs) in the Internet make communication between nodes difficult. Traditional gossip-based Peer Sampling Services (PSS) break down when a high percentage of nodes are behind NATs. I solved this problem by presenting two NAT-aware PSSs: Gozar [14] and Croupier [11]. The nodes in Gozar use one-hop relaying to communicate with the nodes behind NATs, while Croupier implements a gossip-based PSS, but without the use of relaying.

  A few P2P systems tackled the NAT problem by employing third party nodes to establish a connection towards nodes behind NAT, and these may become bottlenecks, menacing the health of the entire system. To address this problem, we presented NATCloud [5], a cloud-assisted NAT-traversal service, where rented cloud resources are added on demand to the overlay, as third party nodes, to help other nodes to make connections to nodes behind NAT.

- **Scalable Publish/Subscribe Systems (2010 - 2012).** P2P overlay networks are attractive solutions for building large-scale publish/subscribe systems. However, scalability comes with a cost: a message published on a certain topic often needs to traverse a large number of

uninterested nodes before reaching all its subscribers. To overcome this problem, we introduced Vitis [15], a gossip-based topic-based publish/subscribe system that significantly decreases the number of relay messages, and scales to an unbounded number of nodes and topics. This is achieved by the novel approach of enabling rendezvous routing on unstructured overlays. We construct a hybrid system by injecting structure into an otherwise unstructured network. The resulting structure resembles a navigable small-world network, which spans along clusters of nodes that have similar subscriptions. The properties of such an overlay make it an ideal platform for efficient data dissemination in large-scale systems.

- **Gossip-based Distribution Estimation (2010 - 2012).** Monitoring the global state of an overlay network is vital for the self-management of P2P systems. Gossip-based aggregation is a well-known technique that provides nodes locally with aggregated knowledge about the state of the overlay network. In [8], we present a gossip-based protocol to estimate the global distribution of attribute values stored across a set of nodes of a network. Our algorithm estimates the distribution both efficiently and accurately.

3.1 List of Publications

According to Google Scholar (April 14, 2018): citation 401, h-index 11, i10-index 12. All the papers are available at: `https://payberah.github.io`.

International Journals

[1] Bouguelia, M.R., Nowaczyk, S., **Payberah, A.H.**, *An Adaptive Algorithm for Anomaly and Novelty Detection in Evolving Data Streams.* Data Mining and Knowledge Discovery, Springer, 2018.

[2] Rahimian, F., **Payberah, A.H.**, Girdzijauskas, S., Jelasity, M., and Haridi, S., *A Distributed Algorithm For Large-Scale Graph Partitioning.* ACM Transactions on Autonomous and Adaptive Systems (TAAS), 10(2), 2015.

[3] **Payberah, A.H.**, Dowling, J., Rahimian, F., and Haridi, S., *Distributed Optimization Of P2P Live Streaming Overlays.* The Springer Computing Journal, Special Issue on Extreme Distributed Systems: From Large Scale to Complexity (Computing), 94(8), pp. 621–647, 2012.

International Conferences and Workshops

[4] Peiro, H., **Payberah, A.H.**, Rahimian, F., Vlassov, V., and Haridi, S., *Boosting Vertex-Cut Partitioning For Streaming Graphs.* IEEE International Congress on Big Data (BigData), pp. 1–8, 2016. **Best Paper Award**

[5] Kavalionak, H., **Payberah, A.H.**, Montresor, A., and Dowling, J., *NATCloud: Cloud-Assisted NAT-Traversal Service.* ACM Symposium on Applied Computing (SAC), pp. 508–513, 2016.

[6] Rahimian, F., **Payberah, A.H.**, Girdzijauskas, S., and Haridi, S., *Distributed Vertex-Cut Partitioning.* IFIP International Conference on Distributed Applications and Interoperable Systems (DAIS), pp. 186–200, 2014.

[7] Rahimian, F., **Payberah, A.H.**, Girdzijauskas, S., Jelasity, M., and Haridi, S., *Ja-Be-Ja: A Distributed Algorithm For Balanced Graph Partitioning.* IEEE International Conference on Self-Adaptive and Self-Organizing Systems (SASO), pp. 51–60, 2013. **Best Paper Award**

[8] **Payberah, A.H.**, Kavalionak, H., Montresor, A., Dowling, J., and Haridi, S., *Lightweight Gossip-based Distribution Estimation.* IEEE International Conference on Communications (ICC), pp. 3439–3443, 2013.

[9] Freitag, F., Navarro, L., Khan, A.M., Baig, R., Escrich, P., Jimenez, J., Pietrosemoli, E., Zennaro, M., Vlassov, V., and **Payberah, A.H.**, *Supporting Cloud Deployment in the Guifi.net Community Network.* IEEE Global Information Infrastructure and Networking Symposium (GIIS), pp. 1–3, 2013.

[10] **Payberah, A.H.**, Kavalionak, H., Kumaresan, V., Montresor, A., and Haridi, S., *CLive: Cloud-Assisted P2P Live Streaming.* IEEE International Conference on Peer-to-Peer Computing (P2P), pp. 79–90, 2012.

[11] Dowling, J., and **Payberah, A.H.**, *Shuffling With A Croupier: NAT-Aware Peer-Sampling.* IEEE International Conference on Distributed Computing Systems (ICDCS), pp. 102–111, 2012.

[12] Rahimian, F., Girdzijauskas, S., **Payberah, A.H.**, and Haridi, S., *Subscription Awareness Meets Rendezvous Routing.* IARIA International Conference on Advances in P2P Systems (AP2PS), pp. 1–10, 2012.

[13] **Payberah, A.H.**, Dowling, J., and Haridi, S., *GLive: The Gradient Overlay As A Market Maker For Mesh-Based P2P Live Streaming.* IEEE International Symposium on Parallel and Distributed Computing (ISPDC), pp. 153–162, 2011.

[14] **Payberah, A.H.**, Dowling, J., and Haridi, S., *Gozar: NAT-Friendly Peer Sampling With One-Hop Distributed NAT Traversal.* IFIP International Conference on Distributed Applications and Interoperable Systems (DAIS), pp. 1–14, 2011.

[15] Rahimian, F., Girdzijauskas, S., **Payberah, A.H.**, and Haridi, S., *Vitis: A Gossip-Based Hybrid Overlay For Internet-Scale Publish/Subscribe.* IEEE International Parallel and Distributed Processing Symposium (IPDPS), pp. 746–757, 2011.

[16] Terelius, H., Shi, G., Dowling, J., **Payberah, A.H.**, Gattami, A., and Johansson, K.H., *Converging An Overlay Network To A Gradient Topology.* IEEE Conference on Decision and Control and European Control Conference (CDC-ECC), pp. 7230–7235, 2011.

[17] **Payberah, A.H.**, Dowling, J., Rahimian, F., and Haridi, S., *Sepidar: Incentivized Market-Based P2P Live-Streaming On The Gradient Overlay Network.* IEEE International Symposium on Multimedia (ISM), pp. 1–8, 2010.

[18] **Payberah, A.H.**, Dowling, J., Rahimian, F., and Haridi, S., *gradienTv: Market-based P2P Live Media Streaming On The Gradient Overlay.* IFIP International Conference on Distributed Applications and Interoperable Systems (DAIS), pp. 212–225, 2010.

Books and Book Chapters

[19] **Payberah, A.H.**, Rahimian, F., *Introduction to Big Data*, In Big Data and Software Defined Networks, Taheri, J. Editor, The IET Big Data Book Series, 2017.

[20] **Payberah, A.H.**, Janzadeh, H., and Kazempour, V., *Linux Device Driver Programming*, Ardalan Press, 2006, ISBN: 964-92399-8-7 (In Farsi).

[21] **Payberah, A.H.**, Sarikhani, M., and Nazari, A., *Linux Process Programming and Management*, Ardalan Press, 2006, ISBN: 964-92399-9-5 (In Farsi).

3.2 Grants

- *STREAMLINE: Improving Competitiveness of European Enterprises through Streamlined Analysis of Data at Rest and Data in Motion*, at SICS, from 2015 to 2018. Funded by H2020, ICT-16-2015, big data. Volume: 3.2M Euro. Responsibility: coauthor of the proposal.
- *BIDAF: A Big Data Analytics Framework for a Smart Society*, at SICS, from 2014 to 2019. Funded by KK-stiffen (KKS). Volume: 2.6M Euro. Responsibility: coauthor of the proposal and the leader of a work package.

3.3 Active Participation in National and International Conferences

- Track Co-Chair, IEEE Consumer Communications and Networking Conference (CCNC), 2017
- PC member, National Iranian CSI Computer Conference (CSICC), 2017
- PC member, EAI International Conference on Smart Objects and Technologies for Social Good (GOODTECHS), Open Challenges in Online Social Networks, 2016
- PC member, IEEE International Conference on Ubiquitous Computing and Communications (IUCC), 2014
- PC member, IARIA International Conference on Communications, Computation, Networks and Technologies (INNOV), 2014
- PC member, IFIP International Conference on Distributed Applications and Interoperable Systems (DAIS), 2013
- Organization committee member, EIT ICT Labs summer school on Cloud and Big Data (`http://ictlabs-summer-school.sics.se`), 2015 and 2016.

3.4 Review and Referee Assignments

The average number of review assignments per year from international periodicals is around four papers.

Journals

- IEEE Transactions on Parallel and Distributed Systems (TPDS)
- Elsevier Journal of Parallel and Distributed Computing (JPDC)
- Elsevier Computer Networks (COMNET)
- Elsevier Computer Communications (COMCOM)
- Elsevier Computers and Electrical Engineering (COMPELECENG)
- Elsevier Information Systems (IS)
- Elsevier Future Generation Computer Systems (FGCS)
- Oxford The Computer Journal (COMPJ)

- Springer Computing (COMPD)
- Springer Peer-to-Peer Networking and Applications (PPNA)
- Academy Publisher Journal of Communications (JCM)
- International Journal of Web Intelligence Consortium (WI)
- International Journal of Concurrency and Computation: Practice and Experience (CPE)
- International Journal of Big Data Intelligence (IJBDI)
- International Journal of Internet Technology (JIT)
- International Journal of Algorithms
- Iranian Journal of Electrical and Computer Engineering (IJECE)

Conferences

- IEEE International Symposium on Parallel and Distributed Processing with Applications (ISPA), 2015
- ACM Special Interest Group on Management Of Data (SIGMOD), 2013
- IEEE International Conference on Peer-to-Peer Computing (P2P), 2008 - 2011
- IFIP International Conference on Distributed Applications and Interoperable Systems (DAIS), 2011
- IEEE Consumer Communications and Networking Conference (CCNC), 2010 and 2013

3.5 National and International Prizes

- **Best paper award** for the paper *Boosting Vertex-Cut Partitioning For Streaming Graphs* in IEEE International Congress on Big Data (BigData), 2016.
- **Best paper award** for the paper *Ja-Be-Ja: A Distributed Algorithm For Balanced Graph Partitioning* in IEEE International Conference on Self-Adaptive and Self-Organizing Systems (SASO), 2013.

3.6 Software

- I have been using a wide range of programming languages and data intensive computing platforms for my work, including Java, Scala, Python, C/C++ and Bash as programming languages, and Hadoop, Spark, Flink, Tensorflow, and GraphLab as processing platforms. All my developed softwares are available at: `http://github.com/payberah`.
- SICSIM, a discrete event, flow level simulator for P2P applications in Java, 2008. Available at: `https://www.sics.se/~amir/sicsim`.
- Learnux, a customized Farsi GNU/Linux distribution based on SLAX, that contains self-learning documents about GNU/Linux, 2005.

4 Teaching Effort

4.1 Graduate Level

Data Intensive Computing (2014 - 2016), KTH, Sweden
In this course, I described the critical technology trends that are enabling cloud computing and the services and applications they offer. The course covers a wide variety of advanced topics in data intensive computing, including distributed file systems, NoSQL databases, processing of different data models, e.g., data-at-rest (batch processing), data-in-motion (stream processing), linked data (graph processing) and structured data, and resource management. The course is mainly based on research papers and Spark ecosystem. The grading of the course is based on several parts: reading assignments, lab assignments, a final project, a presentation, a mid-term exam and a final exam. This course was given as a part of the Master of data science program at KTH in autumn 2016. The course information is available at `https://www.sics.se/~amir/id2221`. The first version of the course was given for PhD students of KTH, SICS employees, and a few from industry, e.g., Ericsson in 2014. Later in Fall 2014, a more enhanced version of the course was presented at AUT in Iran for graduate students. All the information of the course, including the scheduling, materials, and discussion forum are available online `https://www.sics.se/~amir/cloud14`. Moreover, I have been teaching this course as an electronic distance course for students at AUT through a virtual environment prepared by AUT. I received extremely good feedback for this course. The course evaluation results are available online at KTH site.

Distributed Systems (2015), AUT, Iran
This course is an introductory course in distributed systems. The emphasis is on the techniques for creating functional, usable, and high-performance distributed systems. The goals of this course are two fold: (i) to have an understanding of the principles and techniques behind the design of distributed systems, such as concurrency, scheduling, and communication across the network, and (ii) gaining practical experience in designing and implementing real distributed systems. I gave this course as an electronic distance course for AUT students in Spring 2015. The course consists of a series of review questions, which are given after each section, as well as a number of programming assignments that cover the main practical challenges in the course.

Distributed Computing, P2P and Grids (2008 - 2012), KTH, Sweden
This course provides students with basic concepts and principles of large-scale dynamic distributed systems and distributed algorithms. The course is taught in seminar style, and several case studies are included. The topics covered by this course include the common concepts of P2P, e.g., DHT, gossip based algorithms and content distribution. The course was led by Prof. Seif Haridi, when I joined the course as a teaching assistant (TA). During my TA time, I redeveloped some of the lectures, e.g., Chord and Kademlia DHTs, bittorrent, content distribution, P2P streaming and NoSQL databases, which are still used in the course. I also designed the course programming assignments based on our new frameworks, e.g., Kompics.

Network Algorithms (2008 - 2009), KTH, Sweden
The course focuses on distributed algorithms that enable key functions in emerging technologies, such as P2P services, networked control systems and network management for next-generation networks. The course was led by Prof. Rolf Stadler, and it was organized around two project assignments, each of which focusing on a specific topic. I joined the course as the TA of the P2P overlay networks project. I gave lectures on DHTs and designed one of the course projects.

4.2 Undergraduate Level

Operating Systems (2014 - 2015), AUT, Iran
This course leads the students to deepened knowledge in designing operating systems. The topics we cover include concepts of operating systems and systems programming, such as multiprogramming systems (e.g., processes, threads, interprocess communication, and synchronization), memory management (e.g., segmentation and paging), storage management and file systems, I/O systems (e.g., device drivers), and security and protection. Each part is explained in both theory and practice. To give a better understanding of the operating systems to students, I also explained each concept separately in the Linux kernel. I designed seven system programming assignments in C for the course, such that each assignment covers a part of the course book. I have used four different books to prepare the content of the course. The course web page contains all the course slides, assignments, and schedule: `https://www.sics.se/~amir/os14`.

Linux Kernel Architecture and System Programming (2001 - 2006), AUT, Iran
This course was a practical course that helped students to work closely with Linux kernel. It was the first Linux course presented at AUT, and I developed the course from scratch. The course consisted of a number of projects, such that in each one students got familiar with part of the Linux kernel. The slides of this course are available here: `https://www.sics.se/~amir/download.htm`.

Linux Device Driver (2004), MetaNet, Iran
The goal of this course was to give a deep understanding of Linux device driver programming. The participants of the course were not only the AUT students, but also many from industry. The slides of the course are available online `https://www.sics.se/~amir/download.htm`. After this course, I translated two Linux books into Farsi: Linux device driver programming [20] and Linux process programming and management [21].

4.3 Academic Supervising Experience

- MSc students
  - Adam Ulhir, DataFlow graphical user interface, KTH, Sweden, 2017
  - Pradeep Peiris, Distributed multi-dimensional cube over Spark, KTH, Sweden, 2017
  - Xinye Fu, Building evolutionary clustering algorithms on Spark, KTH, Sweden, 2017
  - Alexander Ostman, Gossip-based resource management in cloud, KTH, Sweden, 2017
  - Mohsen Hariri and Mahak Memar (joint project), TuxStream: a locality-aware hybrid tree/mesh overlay for P2P live media streaming, KTH, Sweden, 2010
- BSc students
  - Vahid Pouryousef, Coloring massive graphs using GraphX, AUT, Iran, 2016
  - Sina Sheikholeslami, Stream sampling system on Spark streaming, AUT, Iran, 2016
  - Reyhaneh Shahmohammadi, Anomaly detection of streaming data using Storm, AUT, Iran, 2015
  - Ata Mazloomian, Measuring the effectiveness of different vaccination strategies to prevent disease distribution in a P2P network using gossip algorithms, AUT, Iran, 2015

5 Other Assignments

5.1 Workshops and Schools

- One day workshop, the Spark data intensive platform, Sharif ICT Innovation Center, Iran, August 2016.
- Two days workshop, the Spark ecosystem, EIT ICT Labs summer school on Cloud and Big Data, July 2016.
- One day workshop, introduction to big data platforms, SeRC, Sweden, February 2016.
- One day workshop, the Spark data intensive platform, Cafe Bazzar the Android market in Iran, Iran, January 2015.
- One day workshop, data intensive computing, eFarda e-commerce company, Iran, January 2015.
- One day workshop, the Spark data intensive platform, SICS, Sweden, June 2014.

5.2 Talks

- Keynote speaker, the Stratosphere data intensive platform, EIT ICT Labs and Cloudberry Science and Innovation Days, Luleå University of Technology, Sweden, June 2014.
- Webinar on massive graph processing at AUT Big Data Workshop, Iran, April 2017.
- Guest lecturer, introducing to Spark, as a part of the Data Mining, KTH, Sweden, November 2016.
- Guest lecturer, data intensive computing, as a part of the Internet of Things course, Uppsala University, Sweden, February 2015.
- Guest lecturer, DHT, as a part of the Distributed Computing, Peer-to-Peer and Grids course, KTH Royal Institute of Technology, Sweden, March 2015.
- Guest lecturer, NoSQL databases, as a part of the Distributed Computing, Peer-to-Peer and Grids course, KTH Royal Institute of Technology, Sweden, March 2014.
- Webinar on data intensive computing at Linux festival at AUT, Iran, May 2015.
- Talk on cloud computing at AUT, Iran, April 2015.
- Talk on P2P content distribution (BiTtorent and Spotify) at AUT, Iran, November 2014.