Gagan Aggarwal

Office Address:
Gates 470
Stanford, CA 94305
(650) 723-1787
email: `gagan@cs.stanford.edu`
`http://theory.stanford.edu/~gagan`

Home Address:
221 Ayrshire Farm Rd Apt 201
Stanford, CA 94305
(650) 497-9034

Research Interests

Combinatorial and Optimization Algorithms. Privacy and Security. Auctions and Pricing Algorithms. Load Balancing and Scheduling.

Education

Stanford University, *Stanford, CA* *September 2000 - present*
PhD Candidate in Computer Science
Advisor: Professor Rajeev Motwani
Thesis: Publishing Digital Data: Privacy and Pricing

Stanford University, *Stanford, CA* *September 2000 - 2003*
M.S. in Computer Science
Advisor: Professor Rajeev Motwani

Indian Institute of Technology, *New Delhi, India* *July 1996 - May 2000*
Bachelors (B.Tech.) in Computer Science (GPA 9.32/10).
Advisor: Professor Huzur Saran
Project: Congestion Control and Bandwidth Sharing in the Internet

Academic Awards

Stanford Graduate Fellowship (SGF) for the years 2000-2005, Stanford University.

Award for Academic Excellence for the years 1997-98 and 1999-2000, Kailash House, Indian Institute of Technology, Delhi.

Certificate of Merit for several semesters, IIT, Delhi.

Ranked 92 among more than 100,000 students appearing in the Joint Entrance Examination, 1996 for admission to the Indian Institutes of Technology.

National Talent Search Scholarship for 1994-2000, NCERT, India.

Mathematics Talent Search Scholarship in 1994, National Board of Higher Mathematics, India.

Research and Professional Experience

Research Assistant, Stanford University, *Stanford, CA* *2000 to present*
Advisor: Rajeev Motwani

Teaching Assistant, Stanford University, *Stanford, CA* *Fall 2004*
Professor: Rajeev Motwani
Description: Held office hours and graded assignments for "Randomized Algorithms" (CS 365).

Research Intern, Microsoft Research, *Mountain View, CA* *Summer 2004*
Mentor: Jason Hartline
Description: Worked on mechanism design problems, including derandomization of auctions and limited-supply attribute auctions.

Research Intern, HP Labs, *Palo Alto, CA* *Summer 2003*
Mentors: Nina Mishra and Benny Pinkas
Description: Worked on privacy-related problems, including secure computation of the k^{th}-smallest element of a set.

Research Intern, IBM Almaden Research Center, *San Jose, CA* *Summer 2002*
Mentor: Sridhar Rajagopalan
Description: Worked on models for massive dataset computation. Also worked on algorithms for prefetching to improve cache performance.

Teaching Assistant, Stanford University, *Stanford, CA* *Spring 2002*
Professor: Leo Guibas
Description: Taught recitations, held office hours and graded assignments, examinations and programming project for "Introduction to Algorithms" (CS 161).

Research Intern, AT&T Labs, *Florham Park, NJ* *Summer 2001*
Mentors: Suresh Venkatsubramanain and Balachander Krishnamurthy
Description: Worked on clustering of IP addresses in order to identify autonomous domains.

Senior-year Project, IIT, *New Delhi, India* *Fall 1999 and Spring 2000*
Advisor: Huzur Saran
Project: Congestion Control and Bandwidth Sharing in the Internet.

Research Intern, IBM India Research Lab, *New Delhi, India* *Summer 1999*
Mentor: Rajeev Shorey
Project: Development, implementation and testing of an algorithm for estimating the goodput and losses on a link in the Internet.

Junior-year Project, IIT, *New Delhi, India* *Spring 1999*
Advisor: Huzur Saran
Project: Differentiated Services in the Internet.

Summer Project, IIT, *New Delhi, India* *Summer 1997*
Advisor: S.N. Maheshwari
Project: Software development for the Training and Placement Wing at IIT, Delhi.

PUBLICATIONS

CONFERENCE/JOURNAL PAPERS

1. Gagan Aggarwal, Amos Fiat, Andrew Goldberg, Jason Hartline, Nicole Immorlica, Anna Karlin and Madhu Sudan, "Derandomization of Auctions", accepted to *ACM Symposium on Theory of Computing (STOC)*, Baltimore, May 2005.

2. Gagan Aggarwal, Tomas Feder, Krishnaram Kenthapadi, Rajeev Motwani, Rina Panigrahy, Dilys Thomas and An Zhu, "Anonymizing Tables", *International Conference on Database Theory (ICDT)*, Edinburgh, January 2005.

3. Gagan Aggarwal, Tomas Feder, Rajeev Motwani, Rina Panigrahy and An Zhu, "Algorithms for the Database Layout Problem", *International Conference on Database Theory (ICDT)*, Edinburgh, January 2005.

4. Gagan Aggarwal, Mayank Bawa, Prasanna Ganesan, Hector Garcia-Molina, Krishnaram Kenthapadi, Rajeev Motwani, Utkarsh Srivastava, Dilys Thomas and Ying Xu, "Two Can Keep A Secret: A Distributed Architecture for Secure Database Services", *Conference on Innovative Data Systems Research (CIDR)*, Asilomar, January 2005.

5. Gagan Aggarwal, Mayur Datar, Sridhar Rajagopalan and Matthias Ruhl, "On the Streaming Model Augmented with a Sorting Primitive", *Proceedings of IEEE Symposium on Foundations of Computer Science (FOCS)*, Rome, October 2004.

6. Gagan Aggarwal, Mayank Bawa, Prasanna Ganesan, Hector Garcia-Molina, Krishnaram Kenthapadi, Nina Mishra, Rajeev Motwani, Utkarsh Srivastava, Dilys Thomas, Jennifer Widom and Ying Xu, "Enabling Privacy for the Paranoids", *International Conference on Very Large Data Bases (VLDB)*, Toronto, August 2004.

7. Gagan Aggarwal, Tomás Feder, Rajeev Motwani and An Zhu, "Algorithms for Multi-Product Pricing", *Proceedings of International Colloquium on Automata, Languages and Programming (ICALP)*, Turku, Finland, July 2004.

8. Gagan Aggarwal, Mayur Datar, Nina Mishra and Rajeev Motwani, "On Identifying Stable Ways to Configure Systems", *International Conference on Autonomic Computing (ICAC)*, New York, May 2004.

9. Gagan Aggarwal, Nina Mishra and Benny Pinkas, "Privacy-Preserving Computation of the k^{th}-ranked element", *Proceedings of Eurocrypt*, Interlaken, May 2004.

10. Gagan Aggarwal, Michael Goldwasser, Ming-Yang Kao and Robert Schweller, "Complexities for Generalized Models of Self-Assembly", accepted to SIAM Journal of Computing, a preliminary version appeared in *Proceedings of ACM-SIAM Symposium on Discrete Algorithms (SODA)*, New Orleans, January 2004.

11. Gagan Aggarwal, Rajeev Motwani, Devavrat Shah and An Zhu, "Switch Scheduling via Randomized Edge Coloring", *Proceedings of IEEE Symposium on Foundations of Computer Science (FOCS)*, Boston, October 2003.

12. Gagan Aggarwal, Rajeev Motwani and An Zhu, “The Load Rebalancing Problem”, accepted to Journal of Algorithms, a preliminary version appeared in *Proceedings of ACM Symposium on Parallel Algorithms and Architectures (SPAA)*, San Diego, June 2003.

Submitted Papers

13. Gagan Aggarwal, Ashish Goel and Rajeev Motwani, “Truthful Auctions for Targeted Web Advertisements”, 2005.
14. Gagan Aggarwal and Jason Hartline, “Knapsack Auctions”, 2005.
15. Gagan Aggarwal, Tomas Feder, Krishnaram Kenthapadi, Rina Panigrahy, Dilys Thomas and An Zhu, “Clustering for Anonymity Applications”, 2005.
16. Gagan Aggarwal, Moses Charikar, Xiaosong Ma, Subramanyam Mallela, Dharmendra Modha and Matthias Ruhl, “On Caching in the Demand Prepaging Model”, 2005.

Talks

- “Sampling for Efficient IP Clustering”, AT&T Research Labs, Florham Park, NJ, September 14, 2001.
- “Occupancy Problems with Feedback”, Theory Seminar, Stanford University, Stanford, CA, October 24, 2002.
- “The Load Rebalancing Problem”.
  - IBM Almaden Research Center, San Jose, CA, September 13, 2002.
  - ACM Symposium on Parallel Algorithms and Architectures (SPAA), San Diego, CA, June 9, 2003.
- “Self-Assembly of Squares and Other Shapes”, Qualifying Examination Talk, Stanford University, Stanford, CA, May 27, 2003.
- “Privacy-preserving Clustering”, HP Labs, Palo Alto, CA, July 29, 2003.
- “Privacy-preserving Computation”, HP Labs, Palo Alto, CA, September 18, 2003.
- “Switch Scheduling via Randomized Edge-Coloring”, Bay Area Theory Symposium (BATS), Microsoft Research Silicon Valley Center, Mountain View, CA, January 23, 2004.
- “Secure Computation of the k^{th}-ranked Element”.
  - DIMACS/PORTIA Workshop on Privacy-Preserving Data Mining, Rutgers Univerity, Piscataway, NJ, March 15, 2004.
  - Theory Seminar, Stanford University, Stanford, CA, April 15, 2004.
  - EuroCrypt, Interlaken, Switzerland, May 3, 2004.
  - Security Forum, Stanford University, Stanford, CA, May 19, 2004.
  - Intel Research Lab, Berkeley, CA, August 5, 2004.

- "On Identifying Stable Ways to Configure Systems", International Conference on Autonomic Computing (ICAC), New York, May 18, 2004.
- "Algorithms for Multi-Product Pricing", International Colloquium on Automata, Languages and Programming (ICALP), Turku, Finland, July 14, 2004.
- "Knapsack Auctions".
  - Microsoft Research Silicon Valley Center, Mountain View, CA, September 9, 2004.
  - Theory Seminar, Stanford University, Stanford, CA, October 28, 2004.

Refereeing Work

SIAM Journal of Computing (SICOMP), ACM Symposium on Theory of Computation (STOC), Symposium on Theoretical Aspects of Computer Science (STACS), Computing and Combinatorics Conference (COCOON), Foundations of Software Technology and Theoretical Computer Science (FSTTCS) , Combinatorial and Algorithmic Aspects of Networking (CAAN), ACM SIGMOD International Conference on Management of Data (SIGMOD), ACM SIGMOD Principles of Database Systems (PODS), ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), and International Conference on Machine Learning (ICML).

Relevant Coursework

Algorithms and Data Structures, Algorithm design and Analysis, Advanced Algorithms I and II, Randomized Algorithms, Optimization and Algorithmic Paradigms, Geometric Algorithms, Cryptography and Computer Security, Advanced Cryptography, Logic, Programming Languages, Complexity Theory, Computer Graphics, Artificial Intelligence, Computer Architecture, Operating Systems, Microprocessors, Computer Networks, Advanced Computer Networks, Numerical Analysis, Probability Theory.

REFERENCES

Professor Rajeev Motwani	Stanford University Computer Science Department Room 474 353 Serra Mall, Stanford, CA 94305 (650) 723-6045 `rajeev@cs.stanford.edu`
Professor Dan Boneh	Stanford University Computer Science Department Room 476 353 Serra Mall, Stanford, CA 94305 (650) 725-3897 `dabo@cs.stanford.edu`
Dr. Nina Mishra	HP Labs 1501 Page Mill Road Palo Alto, CA (650) 236-2863 `nmishra@theory.stanford.edu`
Dr. Jason Hartline	Microsoft Research 1065 La Avenida Mountain View, CA (650) 693-2266 `hartline@microsoft.com`

DESCRIPTIONS OF PAPERS

1. Gagan Aggarwal, Amos Fiat, Andrew Goldberg, Jason Hartline, Nicole Immorlica, Anna Karlin and Madhu Sudan, "Derandomization of Auctions", accepted to *ACM Symposium on Theory of Computing (STOC)*, Baltimore, May 2005.

   We study truthful auctions that seek to maximize profit. It has been shown that there does not exist any symmetric, deterministic, truthful mechanism that achieves any finite approximation ratio. We show the existence of a deterministic auction that achieves a constant factor of the optimal profit (with a small additive loss). Our technique uses asymmetry to bypass the symmetric impossibility result and shows that asymmetry is essentially as powerful as randomization in a certain context of mechanism design. We give a fairly general derandomization technique for turning any randomized mechanism into an asymmetric deterministic one with approximately the same properties. Our construction involves solving an exponential-sized flow problem and thus is not polynomial-time computable. However, we show that for an interesting problem related to auctions, the 2-color *color-guessing problem*, this derandomization can indeed be performed in polynomial time.

2. Gagan Aggarwal, Tomas Feder, Krishnaram Kenthapadi, Rajeev Motwani, Rina Panigrahy, Dilys Thomas and An Zhu, "Anonymizing Tables", *International Conference on Database Theory (ICDT)*, Edinburgh, January 2005.

   We consider the problem of releasing tables from a relational database containing personal records, while ensuring individual privacy and maintaining data integrity to the extent possible. One of the techniques proposed in the literature is *k-anonymization*. The idea is to hide table entries such that every individual is lost in a crowd of k people. More precisely,a release is considered *k-anonymous* if the information corresponding to a person contained in the release cannot be distinguished from at least $k-1$ other persons whose information also appears in the release. We study the *k-Anonymity* optimization problem where the goal is to hide as few table entries as possible while ensuring that the released version is k-anonymous. We show that the k-Anonymity problem is NP-hard even when the attribute values are ternary. On the positive side, we provide an $O(k)$-approximation algorithm for the problem. This improves upon the previous best known approximation factor of $O(k \log k)$.

3. Gagan Aggarwal, Tomas Feder, Rajeev Motwani, Rina Panigrahy and An Zhu, "Algorithms for the Database Layout Problem", *International Conference on Database Theory (ICDT)*, Edinburgh, January 2005.

   We present a formal analysis of the database layout problem, i.e., the problem of determining how database objects such as tables and indexes are assigned to disk drives. Optimizing this layout has a direct impact on the I/O performance of the entire system. The traditional approach of striping each object across all available disk drives is aimed at optimizing I/O parallelism; however, it is suboptimal when queries co-access two or more database objects, e.g., during a merge join of two tables, due to the increase in random disk seeks. In fact, the optimization problem is intractable in general. We employ techniques from approximation algorithms to present provable performance guarantees. We show that while optimally exploiting I/O parallelism alone suggests uniformly striping data objects (even for heterogeneous files and disks), optimizing random disk access alone would assign each data object to a single disk drive. This confirms the intuition that the two effects are in tension with each other. We provide approximation algorithms in an attempt to optimize the trade-off between the two effects. We show that our algorithm achieves the best possible approximation ratio.

4. Gagan Aggarwal, Mayank Bawa, Prasanna Ganesan, Hector Garcia-Molina, Krishnaram Kenthapadi, Rajeev Motwani, Utkarsh Srivastava, Dilys Thomas and Ying Xu, “Two Can Keep A Secret: A Distributed Architecture for Secure Database Services”, *Conference on Innovative Data Systems Research (CIDR)*, Asilomar, January 2005.

   Recent trends towards database outsourcing, as well as concerns and laws governing data privacy, have led to a great interest in enabling secure database services. Previous approaches to enabling such a service have been based on data encryption, causing a large overhead in query processing. We propose a new, distributed architecture that allows an organization to outsource its data management to *two* untrusted servers while preserving data privacy. We show how the use of two non-colluding servers enables efficient partitioning of data so that the contents at any one server are guaranteed not to breach data privacy. We show how to optimize and execute queries in this architecture and discuss new challenges that emerge in designing the database schema.

5. Gagan Aggarwal, Mayur Datar, Sridhar Rajagopalan and Matthias Ruhl, “On the Streaming Model Augmented with a Sorting Primitive”, *Proceedings of IEEE Symposium on Foundations of Computer Science (FOCS)*, Rome, October 2004.

   The need to deal with massive data sets in many practical applications has led to a growing interest in computational models appropriate for large inputs. The most important quality of a realistic model is that it can be efficiently implemented across a wide range of platforms and operating systems. In this paper, we study the computational model that results if the streaming model is augmented with a sorting primitive. We argue that this model is highly practical, and that a wide range of important problems can be efficiently solved in this (relatively weak) model. Examples are undirected connectivity, minimum spanning trees, and red-blue line segment intersection, among others. This suggests that using more powerful, harder to implement models may not always be justified. We also show a hardness result for the “streaming and sorting” model, which demonstrates that the main limitation of this model is that it can only access one data stream at a time. Finally, we compare our model to a popular restriction of external memory algorithms that access their data mostly sequentially.

6. Gagan Aggarwal, Mayank Bawa, Prasanna Ganesan, Hector Garcia-Molina, Krishnaram Kenthapadi, Nina Mishra, Rajeev Motwani, Utkarsh Srivastava, Dilys Thomas, Jennifer Widom and Ying Xu, “Enabling Privacy for the Paranoids”, *International Conference on Very Large Data Bases (VLDB)*, Toronto, August 2004.

   P3P is a set of standards that allow corporations to declare their privacy policies. Hippocratic Databases have been proposed to implement such policies within a corporation’s datastore. From an end-user individual’s point of view, both of these rest on an uncomfortable philosophy of trusting corporations to protect his/her privacy. Recent history chronicles several episodes when such trust has been willingly or accidentally violated by corporations facing bankruptcy courts, civil subpoenas or lucrative mergers. We contend that data management solutions for information privacy must restore controls in the individual’s hands. We suggest that enabling such control will require a radical re-think on modeling, release, and management of personal data.

7. Gagan Aggarwal, Tomás Feder, Rajeev Motwani and An Zhu, “Algorithms for Multi-Product Pricing”, *International Colloquium on Automata, Languages and Programming (ICALP)*, Turku, Finland, July 2004.

   The availability of data on consumer profiles collected e.g. by websites has opened new possibilities for companies to increase their revenue via data mining techniques. One approach is to strategically set prices of various products, taking into account the profiles of consumers. We study algorithms for the multi-product pricing problem, whereby, given consumer preferences among products, their budgets, and the costs of production, the goal is to set prices of multiple products from a single company, so as to maximize the overall revenue of the company. We present approximation algorithms as well as negative results for several variants of the multi-product pricing problem, modeling different purchasing patterns and market assumptions.

8. Gagan Aggarwal, Mayur Datar, Nina Mishra and Rajeev Motwani, "On Identifying Stable Ways to Configure Systems", *International Conference on Autonomic Computing (ICAC)*, New York, May 2004.

   We consider the often error-prone process of initially building and/or reconfiguring a computer system. We formulate a new optimization framework for capturing certain aspects of this system (re)configuration process. We describe offline and online algorithms that could aid operators in making decisions for how best to take actions on their computers so as to maintain the health of their systems.

9. Gagan Aggarwal, Nina Mishra and Benny Pinkas, "Privacy-Preserving Computation of the k^{th}-ranked element", *Proceedings of Eurocrypt*, Interlaken, May 2004.

   Given two or more parties possessing large, confidential datasets, we consider the problem of securely computing the k^{th}-ranked element of the union of the datasets, e.g. the median of the values in the datasets. We investigate protocols with sublinear computation and communication costs. In the two-party case, we show that the k^{th}-ranked element can be computed in $\log k$ rounds, where the computation and communication costs of each round are $O(\log M)$, where $\log M$ is the number of bits needed to describe each element of the input data. The protocol can be modified to provide efficient security against a malicious adversary without the use of commitments. In the multi-party setting, we show that the k^{th}-ranked element can be computed in $\log M$ rounds, with $O(s \log M)$ overhead per round, where s is the number of parties. The multi-party protocol can be used in the two-party case and can also be made secure against a malicious adversary efficiently. Our protocols achieve an upper bound not far from the communication complexity lower bound of $\Omega(\log M)$.

10. Gagan Aggarwal, Michael Goldwasser, Ming-Yang Kao and Robert Schweller, "Complexities for Generalized Models of Self-Assembly", accepted to SIAM Journal of Computing, a preliminary version appeared in *Proceedings of ACM-SIAM Symposium on Discrete Algorithms (SODA)*, New Orleans, January 2004.

    We study the complexity of tile self-assembly. There exists a tight lower bound of $\Omega(\frac{\log N}{\log\log N})$ on the tile complexity of assembling an $N \times N$ square for almost all N. We consider the problem of reducing the tile complexity for self-assembly through several natural generalizations of the model. One of our results is a tile set of size $O(\sqrt{\log N})$ which assembles an $N \times N$ square in a model which allows flexible glue strength between non-equal glues. This result is matched by a lower bound dictated by Kolmogorov complexity. For three other generalizations, we show that the $\Omega(\frac{\log N}{\log\log N})$ lower bound applies to $N \times N$ squares. At the same time, we demonstrate that there are some other shapes for which these generalizations allow reduced tile sets. Specifically, for thin rectangles with length N and width k, we provide a tighter lower bound of $\Omega(\frac{N^{\frac{1}{k}}}{k})$ for the standard model, yet we also give a construction which achieves $O(\frac{\log N}{\log\log N})$ complexity in a model in which the temperature of the tile system is adjusted during assembly. We also investigate the problem of verifying whether a given tile system uniquely assembles into a given shape, and show that this problem is NP-hard.

11. Gagan Aggarwal, Rajeev Motwani, Devavrat Shah and An Zhu, "Switch Scheduling via Randomized Edge Coloring", *Proceedings of IEEE Symposium on Foundations of Computer Science (FOCS)*, Boston, October 2003.

    The main component of an internet router is an $n \times n$ switch which routes packets from input to output ports. Such a switch can be viewed as a bipartite graph with the input and output ports as the two vertex sets. Packets arriving at input port i and destined for output port j can be modeled as an edge from i to j. Current switch scheduling algorithms view the routing of packets at each time step as a selection of a bipartite matching. We take the view that the switch scheduling problem across a *sequence* of time-steps is an instance of the edge coloring problem for a bipartite multigraph. Implementation considerations lead us to seek edge coloring algorithms for bipartite multigraphs that are *fast, decentralized,* and *online*. We present a randomized algorithm which has the desired properties, and uses only a near-optimal $\Delta + o(\Delta)$ colors on *dense* bipartite graphs arising in the context of switch scheduling. This algorithm extends to non-bipartite graphs as well. It leads to a novel switch scheduling algorithm which, for stochastic online edge arrivals, is *stable*, i.e., the queue length at each input port is bounded at all times. This is the first *decentralized* switch scheduling algorithm that is also guaranteed to be stable.

12. Gagan Aggarwal, Rajeev Motwani and An Zhu, "The Load Rebalancing Problem", accepted to Journal of Algorithms, a preliminary version appeared in *Proceedings of ACM Symposium on Parallel Algorithms and Architectures (SPAA)*, San Diego, June 2003.

    In the classical load balancing or multiprocessor scheduling problem, we are given a sequence of jobs of varying sizes and are asked to assign each job to one of the m empty processors. A typical objective is to minimize makespan, the load on the heaviest loaded processor. Since in most real world scenarios the load is a dynamic measure, the initial assignment may be not remain optimal with time. Motivated by such considerations in a variety of systems, we formulate the problem of *load rebalancing* — given a possibly suboptimal assignment of jobs to processors, relocate a set of the jobs so as to decrease the makespan. Specifically, the goal is to achieve the best possible makespan under the constraint that no more than k jobs are relocated. We also consider a generalization of this problem where there is an arbitrary cost function associated with each job relocation. Since the problem is clearly NP-hard, we focus on approximation algorithms. We construct a sophisticated algorithm which achieves a 1.5-approximation, with near linear running time. We also show that the problem has a PTAS, resolving the complexity issue.

13. Gagan Aggarwal and Jason Hartline, "Knapsack Auctions", Manuscript, 2005.

    We study the problem of selling a limited amount of divisible good. Each agent wants a publically-known amount of the good. The goal is to design an auction that uses the knowledge of the demand of each bidder in order to get a higher revenue than the no-information case. We compare the revenue of the auction to the optimal monotone pricing revenue, whereby a bidder with lower demand cannot be charged a higher price. Let n be the number of people who get the good under optimal monotone pricing. We present an auction that achieves a constant fraction of the optimal monotone pricing revenue with an additive loss equal to $\log\log\log n$ times the value of the highest bid.

14. Gagan Aggarwal, Tomas Feder, Krishnaram Kenthapadi, Rina Panigrahy, Dilys Thomas and An Zhu, "Clustering for Anonymity Applications", Manuscript, 2005.

    We study the problem of clustering points in a metric space with the objective of minimizing the *cellular* cost. The cellular cost of a clustering is defined as the sum of the setup cost of open cluster centers and the service cost, with each point paying a service cost equal to the distance of its cluster center from the farthest point in the cluster. This measure arises naturally in several application areas including anonymous cellular networks. With this measure, the assignment of a point to a cluster can increase the service cost of all members of the cluster. We develop new techniques to handle this issue, and present a primal-dual algorithm that achieves a 4-approximation for this problem. We then consider an additional constraint which specifies that each cluster should have at least r points assigned to it. This is motivated by privacy (anonymity) considerations. With this constraint, we study the cellular and k-center objective functions, and present constant-factor approximations for both. Finally, we relax the clustering problem by allowing the solution to leave at most an ε fraction of the points unclustered. We study this relaxation along with the minimum cluster size constraint for the cellular and k-center objective functions, and present constant-factor approximation algorithms for both.

15. Gagan Aggarwal, Moses Charikar, Xiaosong Ma, Subramanyam Mallela, Dharmendra Modha and Matthias Ruhl, "On Caching in the Demand Prepaging Model", Manuscript, 2005.

    Belady's MIN is the optimal offline algorithm for caching in a demand paging environment. We study a *demand prepaging* model that allows sequential prefetching of adjacent pages upon misses. Temam proposed a supposedly optimal offline algorithm in this setting, called Extended Belady (X-BELADY). By giving a counter example, we show that X-BELADY is not optimal. We show however, that X-BELADY is within a factor of 2 of optimal.