

## RESEARCH EXPERIENCES

### Apollo: Big Data Fact Finder

Mar 2012 – Current

**Project Website:** <http://apollo.cs.illinois.edu>

Apollo deals with the quality issues of data fusion on streaming big data. It is a newly developed platform for deriving facts from participatory sensing data. Present implementation of Apollo takes crowd-sourced data from social networks like Twitter, and applies various information processing tools to assess credibility, recognize events and collaborations, and answer queries. However, this implementation is not scalable, as (1) it runs on a single cpu, and (2) the computation is not online and operates on a whole at set at a time. In this project, we seek to:

- (1) Explore the possible scopes for scalability and develop a scalable version.
- (2) Derive query coverage to support multiple queries simultaneously.
- (3) Implement a real time and online version of Apollo.

### Diversifying Source Selection in Social Sensing

Mar 2012 – Apr 2012

**Publication:** Md Yusuf Sarwar Uddin, **Md Tanvir Al Amin**, Hieu Le, Tarek Abdelzaher, Boleslaw Szhymancki, Tommy Nguyen, "On Diversifying Source Selection in Social Sensing," **International Conference on Networked Sensing Systems (INSS) 2012**, Belgium.

**Available at:** <http://web.engr.illinois.edu/~maamin2/papers/uddin-inss12.pdf>

**Abstract:** In this project we develop algorithms for improved source selection in social sensing applications that exploit social networks (such as Twitter, Flickr, or other mass dissemination networks) for reporting. The collection point in these applications would simply be authorized to view relevant information from participating clients (either by explicit client-side action or by default such as on Twitter). Social networks, therefore, create unprecedented opportunities for the development of sensing applications, where humans act as sensors or sensor operators, simply by posting their observations or measurements on the shared medium. Resulting social sensing applications, for example, can report traffic speed based on GPS data shared by drivers, or determine damage in the aftermath of a natural disaster based on eye-witness reports. A key problem, when dealing with human sources on social media, is the difficulty in ensuring independence of measurements, making it harder to distinguish fact from rumor. This is because observations posted by one source are available to its neighbors in the social network, who may, in-turn, propagate those observations without verifying their correctness, thus creating correlations and bias. A corner-stone of successful social sensing is therefore to ensure an unbiased sampling of sources that minimizes dependence between them. This paper explores the merits of such diversification. It shows that a diversified sampling is advantageous not only in terms of reducing the number of samples but also in improving our ability to correctly estimate the accuracy of data in social sensing.

### PhotoNet

Sep 2011 – Jan 2012

**Publication:** Md Yusuf Sarwar Uddin, **Md Tanvir Al Amin**, Tarek Abdelzaher, Arun Iyengar, Ramesh Govindan, "PhotoNet+: Outlier-resilient Coverage Maximization in Visual Sensing Applications," **ACM/IEEE Information Processing in Sensor Networks (IPSN) 2012**, China.

**Available at:** <http://web.engr.illinois.edu/~maamin2/papers/uddin-ipsn12.pdf>

Md Yusuf Uddin, **Md Tanvir Al Amin**, and Tarek Abdelzaher, Arun Iyengar, Ramesh Govindan, "On Achieving Diversity in the Presence of Outliers in Participatory Camera Sensor Networks," UIUC Technical Report, 2012

**Available at:** <http://web.engr.illinois.edu/~maamin2/papers/uddin-techrep12.pdf>

**Abstract:** This project addresses the problem of collection and delivery of a representative subset of pictures, in participatory camera networks, to maximize coverage when a significant portion of the pictures may be redundant or irrelevant. Consider, for example, a rescue mission where volunteers and survivors of a large-scale disaster scout a wide area to capture pictures of damage in distressed neighborhoods, using handheld cameras, and report them to a rescue station. In this participatory camera network, a significant amount of pictures may be redundant (i.e., similar pictures may be reported by many) or irrelevant (i.e., may not document an event of interest). Given this pool of pictures, we aim to build a protocol to store and deliver a smaller subset of pictures, among all those taken, that minimizes redundancy and eliminates irrelevant objects and outliers. While previous work addressed removal of redundancy alone, doing so in the presence of outliers is tricky, because outliers, by their very nature, are different from other objects, causing redundancy minimizing algorithms to favor their inclusion, which is at odds with the goal of finding a representative subset. To eliminate both outliers and redundancy at the same time, two seemingly opposite objectives must be met together. The contribution of this paper lies in a new prioritization technique (and its in-network implementation) that minimizes redundancy among delivered pictures, while also reducing outliers.

### Speeding Up the Web

Aug 2011 – Feb 2012

**Report:** Md Tanvir Al Amin, Fatemeh Saremi, Philip Brighten Godfrey, "On Prioritizing Short Flows,"

**Available at:** <http://web.engr.illinois.edu/~maamin2/papers/amin-report12.pdf>

**Abstract:** User engagement in consumer Internet today mostly deals with web applications. These applications typically result in a traffic pattern consisting of mostly short flows, flows that have less than 400 packets to send. As the latest Internet applications like Software as a Service, Computing as a Service, and others become more and more ubiquitous, total Internet traffic will be more skewed towards short flows. But, congestion control algorithms like TCP and XCP are particularly not suited to short flows, and RCP has still room for improvement in this domain. In this project, we have at first done an analysis of the flow completion time under ideal processor sharing model. Our findings suggest that the best way to improve the performance of the tiny flows that have very small number of packets to send is either improving the round trip delay or devising some protocol that does not wait for SYN-ACK and starts sending data immediately. However, we can improve the flows for which data transmission time is comparable to the RTT. Based on our findings, we have proposed a prioritized congestion control scheme that particularly improves user experience in the Internet by prioritizing these specific type of short flows. We have simulated our protocol using *ns-2* using both core-router and edge-router traffic models. Simulation results show faster completion time for short flows and improved total completion times for all flows, confirming the effectiveness of this idea.

### An Efficient Multi-hop Sybil-resilient DHT

Apr 2010 – Apr 2011

**Publication:** Md Tanvir Al Amin, Reaz Ahmed, "Socially Influenced Hierarchy and Admission Control Mechanism for Sybil Resilience in Distributed Hash Tables," 2012. (*In preparation*)

**Presentation & Talks:** Sybil Attack in P2P Networks, Survey on Sybil Defense Mechanisms

**Available at:** [http://tanviramin.com/documents/Project\\_Survey\\_Sybil.ppt](http://tanviramin.com/documents/Project_Survey_Sybil.ppt)

**Abstract:** DHTs (Distributed Hash Tables) are data structures over an overlay network, implementing efficient insert and lookup of key-value pairs in a decentralized manner. DHTs have been subject to the well-known Sybil attack, where an adversary creates many fake identities (known as Sybil nodes) in order to increase its influence and deny service to the honest participants. There are heuristic algorithms to detect Sybil nodes using trust relationships in the social networks. But a Sybil-resilient DHT with scalable routing and table size is still missing. We design the missing system in this project.

We exploit the power law degree distributions, and cohesive community nature of social networks along with the already used fast mixing properties. Random walks are used to identify members of a social cluster in a distributed manner. Each cluster elects a leader, which bootstraps a DHT with logarithmic table size for that cluster. A DHT will deny admission to the users not connected to the community in a cohesive manner. Thus, this is an admission control mechanism, which classifies nodes based on the attachment of a node to a social cluster. As the number of attack edges between Sybil nodes and an “honest community” is low, sybil nodes are effectively prevented from joining the DHT and subverting the routing procedure. As there is a different DHT for each social cluster, the routing is locality-preserving, resulting in an “organic routing” through socially trusted nodes. On the other hand, leaders of each cluster will communicate to one another, and they will form a multicast group to co-ordinate in inter-cluster routing and key-space distribution. Other solutions either increase the finger table size significantly or increase number of routing hops to provide Sybil-resilience; our proposed scheme successfully keeps both of these crucial performance parameters in control. It treats each community of a social network independently. While other sybil-resistant routing schemes just work between one honest region and one Sybil region, our protocol is supposed to be capable to work in the more practical case where there are multiple communities. Moreover, this method neither assumes any unrealistically large table size, nor it makes the routes suboptimal. Simulation results follow the theoretical findings and show significant improvement over the existing protocols.

### Load Balanced Broadcasting in Wireless Ad Hoc Networks

Jan 2008 – Mar 2009

**Publication:** Md Tanvir Al Amin, Sukarna Barua, Ashikur Rahman, and Shudip Vhaduri, “Load Aware Broadcast in Mobile Ad Hoc Networks,” **IEEE International Conference on Communications (IEEE ICC)**, June 14-18 2009, Dresden, Germany. (**Acceptance rate 35%**)

**Available at:** <http://web.engr.illinois.edu/~maamin2/papers/amin-icc09.pdf>

**Abstract:** The primary performance objective of most of the broadcast schemes in an ad-hoc wireless network is to minimize the total number of packet forwarding while maintain maximum reachability. To achieve this objective, most broadcasting schemes start with flooding and attempt to reduce its extent with the assistance of some greedy heuristics. In this paper we show that, using these greedy heuristics results in a biased load distribution throughout the network. Some nodes become heavily loaded and consequently packets through those nodes, whether unicast or broadcast, experience significantly larger delay. As total necessary packet forwards are not distributed evenly, contention and collision also increase at some regions, while remain relatively less at other regions. As a result the highly loaded nodes also experience low battery life. Thus the resources are not utilized properly. While unicast communication has been studied in this context, little has been done to improve the load balancing of broadcast communication. We address this issue, and propose various methods to evenly distribute the load caused by broadcast packets. Our algorithms take various reactive measures to dynamically include less loaded nodes in the forward list, while maintaining the total number of packet forwards low. Detailed simulation shows fair scheduling of resources and significant improvement in distribution of load, latency and overall performance.

### Intelligent Traffic: Smart Phones, Statistical Learning, and Cloud

Aug 2010 – Mar 2010

**Abstract:** Severe traffic congestion in road networks is a major problem. In this project, our main goal is to evenly balance the traffic in the road network, minimize congestion, and decrease overall delay. Basic idea is to collect data on traffic congestion, feed them to servers backed up by cloud, analyze it using statistical learning algorithms, and process user queries. The first step is achieved using people. People would like to feed updates of traffic condition of a certain place at a certain time, using appropriate applications in cell phones, in exchange of some incentive. Some alternate means of collecting data, like data from wireless sensors installed at intersections, feedback from micro or pico cell towers, or GPS uploaded vehicle position are also used. With the collected data, we are working on an application hosted in cloud, which continually updates a traffic congestion map, and provides the vehicle drivers with the best next hop. The next hop is selected considering the present congestion map, according to criterion like fastest route, shortest path, or minimum cost, and also with an objective to balance the traffic over the road network evenly. Some statistical algorithms are also used

to clean and analyze data, and learn about congestion trends based on time of the day, day of the week, or season. Using this knowledge base, services like prediction of total delay in a route at a certain time, best means of transport to a destination, road conditions, etc are also provided.

### DoS Attack by Channel Jamming in Wireless Networks

Nov 2009 – Apr 2010

**Report:** Md Tanvir Al Amin, Sukarna Barua, Nashid Shahriar, Ashikur Rahman, "Maximizing Interruption in Wireless Networks: An Intruder's Perspective," 2010

**Abstract:** Due to the inherent broadcasting nature, wireless networks are vulnerable to several kinds of security attacks like overhearing, malicious association, denial of service and so on. In this project, we have discussed an important security problem, which we call jamming. Here, intruders try to compromise network operations by strategically placing nodes physically interfering with the wireless channel. The problem has some similarity with coverage problems in wireless sensor networks, though it is actually the inverse. Theoretically the problem is NP-Hard. Using some results of Euclidean geometry and graph theory, we have achieved some bounds on how easy it may be for an intruder to launch a DoS attack or a sniffer to overhear all the communication going on with minimum power budget. Using our results, we have developed a heuristic algorithm to jam a wireless network with minimum number of nodes. We also have given directions on how to better place the wireless nodes to make it harder to jam completely.

### Wireless Sensor Network based Surveillance System

Jan 2008 – Jun 2008

**Report:** Md Tanvir Al Amin, Mahmudul Islam, Ahmed Khurshid "Low Cost Surveillance using Wireless Sensor Networks," 2010

**Project Page:** <http://sites.google.com/site/wnsurv>

**Abstract:** In this project, infra red sensors are used as intrusion detectors. Several such sensors, equipped with microcontroller, radio transmitter and receiver unit, are scattered throughout an area, forming a wireless sensor network. Each has a unique address. They communicate over a shared channel using an on board MAC protocol. When a security breach occurs at any sensor, the base station is notified. There is a mounted camera on the base station, which rotates and focuses to the suspected point and records the situation. The exact position of each sensor and the camera controller unit can be calibrated by user using either a web interface or desktop software. Strength of the system is that: the motes, the sensor units and the camera driver, everything is engineered from the scratch, using inexpensive and easily available components. Hence it yields in a low cost surveillance solution.

### Reducing Noise from Natural Images using Algebraic Methods

Jun 2006 – Aug 2006

**Publication:** Md Tanvir Al Amin and Anupam Bhattacharjee, "Impulsive Noise Reduction in Natural images by Plane and Paraboloid Regression," 9th International Conference on Computers and Information Technology (ICIT), Bangladesh, December 2006 (**Acceptance rate 25%**, 117 / 465)

- [Extended Version] Md Tanvir Al Amin, "Reduction of Impulsive Noise in Continuous-Tone Images by Regression Analysis," IEEE Bangladesh Branch Student Paper Contest 2007 (**3<sup>rd</sup> Prize**)

**Available at:** <http://web.engr.illinois.edu/~maamin2/papers/amin-iccit06.pdf>

**Presentation:** [http://tanviramin.com/documents/amin\\_noisereduction\\_iccit06.ppt](http://tanviramin.com/documents/amin_noisereduction_iccit06.ppt)

**Abstract:** Obtaining a clear and lucid image by reducing the noise to a minimal level is one of the most fundamental research topics in image processing. Necessity for reducing noise is not only for aesthetic purposes, but also due to its vital role in the success of image processing and image understanding

algorithms. As there is no way for total elimination, several methods have been developed, depending on the type of noise. In this paper, we devise a method for reducing impulsive noise and study its noise detection and image restoration performance. A sweeping window of certain dimension calculates how well a plane can be fitted over the pixels currently inside the window and then examines each pixel according to some rules whether it is part of noise or part of signal. Final decision about each pixel is taken by counting the majority verdicts about it. Fitting a general paraboloid equation only through the uncorrupted pixels in each window and taking the mean of all suggested values find values of the corrupted pixels. The method can be applied to a vast area of real world applications like digital photography, medical imaging, computer vision and so on. Simulation reveals that on average, our method has success rate of nearly 97% in case of low noise density and nearly 93% in case of medium noise density.

### Architecture for a Multichannel Internet TV System

Jan 2005 – Jul 2005

**Publication:** Md Tanvir Al Amin, Imranul Hoque, and Md Yusuf Sarwar Uddin, "A Service Oriented Architecture for a Multi Channel Internet TV System," 8th International Conference on Computers and Information Technology (ICIT), Bangladesh, 2005. (**Acceptance rate 31%**, 248 / 807)

**Available at:** <http://web.engr.illinois.edu/~maamin2/papers/amin-iccit05.pdf>

**Presentation:** [http://tanviramin.com/documents/amin\\_internettv\\_iccit05.ppt](http://tanviramin.com/documents/amin_internettv_iccit05.ppt)

**Abstract:** Transmitting TV programs over the Internet have always been a great issue in multimedia service. With the advent of broadband internet connectivity, real time video and audio delivery on the Internet is getting popular day by day. Although the industry is making great and highly publicized plans for future (inter) national video on demand, there is lack of robust, scalable and interoperable architecture for Internet television system. In this paper we present such architecture. TV Program Servers, which are capable of capturing and transmitting media, registers with the Broker Server. The Internet TV Clients query the Broker Server, which maintains a list of available program servers and a complex data structure about the clients already connected. The architecture can be implemented over the existing network and no special expensive hardware setup is required. The application of the service-oriented paradigm makes the registration, composition and discovery of services efficient, and hence the system is perfectly suitable for medium to large scale Internet Television.

### Survey on Multi-core Processors and Programming Environments

Nov 2008 – Dec 2008

**Report:** Md Tanvir Al Amin, Multi-core: Adding a New Dimension to Computing

**Available At:** <http://arxiv.org/abs/1011.3382>

**Abstract:** Sustaining development and advancement in electronics and fabrication techniques has caused the devices to shrink in size and become smaller, paving the quest for increasing density and clock speed. That quest has suddenly come to a halt due to fundamental bounds applied by physical laws. The microprocessor industry has now started exploring the technology along a different dimension: increasing the number of independent processor cores packed in a single package. Such processors, commonly known as multi-core processors has gained much attention from the academia and the industry, and even embedded devices today contain them. In this survey, we explore state of the art technologies for multi-core processors, existing software tools to support parallelism and trend of present and future research in this field, we conclude that next few decades are going to be marked by this "Ubiquitous parallel processing".

### Survey on Probabilistic Models of Genome Assembly

Jan 2010 – Mar 2010

**Abstract:** Algorithms for Genome Assembly have traditionally used Shortest Common Superstring (SCS) as the objective. Recently, this objective has been questioned. The reason is the presence of multiple identical or nearly identical stretches of DNA in the sequence, which are shorter than the read

length. Clearly, these repeats are going to be under-represented by the Shortest Common Substring. Moreover, with the advent of NGS (Next Generation Sequencing) technology, short reads with high coverage are possible. Hence, researchers are now actively exploring the newer avenues of GA research comprising of probabilistic models of likelihood. In this survey, we have explored some of these works by Paul Medvedev, Michael Brudno (Maximum Likelihood Genome Assembly) and by Daniel Zerbino (Velvet Assembler), and have suggested on possible improvements on these methods.

**Presentation & Talks:** Maximum Likelihood Genome Assembly

**Available at:** [http://tanviramin.com/documents/genome\\_assembly.ppt](http://tanviramin.com/documents/genome_assembly.ppt)

### Encryption using Inverse Huffman Tree

Jun 2007 – Oct 2007

Worked with Dr. M. Kaykobad on an analysis of hardness of some special cases of Inverse Huffman Tree problem. We also proposed a new data encryption approach based on Huffman Tree. As the Inverse Huffman Tree problem is NP-Complete, it has some merits in encryption. We analyzed some cases when the problem becomes particularly easy, so breaking the code is easier; and vice versa.

### References

- [1] R. Ganti, N. Pham, H. Ahmadi, S. Nangia, and T. Abdelzaher. GreenGPS: A participatory sensing fuel- efficient maps application. In *Proc. Mobisys*, June 2010.
- [2] M. Y. S. Uddin, M. T. A. Amin, T. Abdelzaher, A. Iyengar, and R. Govindan. On achieving diversity in the presence of outliers in participatory camera sensor networks. Technical report, University of Illinois at Urbana-Champaign (UIUC), 2012.
- [3] M. Y. S. Uddin, M. T. A. Amin, T. Abdelzaher, A. Iyengar, and R. Govindan. Photonet+: Outlier-resilient coverage maximization in visual sensing applications. In *Proc. Information Processing in Sensor Networks (IPSN)*, April 2012.
- [4] M. Y. S. Uddin, M. T. A. Amin, H. Le, T. Abdelzaher, B. Szhymski, and T. Nguyen. On diversifying source selection in social sensing. In *Proc. International Conference on Networked Sensing Systems (INSS)*, June 2012.
- [5] M. Y. S. Uddin, H. Wang, F. Saremi, G.-J. Qi, T. Abdelzaher, and T. Huang. Photonet: A similarity-aware picture delivery service for situation awareness. In *Proc. IEEE Real-time Systems Symposium (RTSS)*, December 2011.
- [6] D. Wang, H. Le, L. Kaplan, and T. Abdelzaher. On truth discovery in social sensing: A maximum likelihood estimation approach. In *Proc. 11th ACM/IEEE Conference on Information Processing in Sensor Networks (IPSN)*, April 2012.
- [7] J. Dean and S. Ghemawat. MapReduce: Simplified data processing on large clusters. In OSDI, pages 137–150, 2004.
- [8] M. Isard, M. Budiu, Y. Yu, A. Birrell, and D. Fetterly. Dryad: Distributed data-parallel programs from sequential building blocks. In EuroSys, pages 59–72, 2007.
- [9] Matei Zaharia, Mosharaf Chowdhury, Michael Franklin, Scott Shenker, Ion Stoica, Spark: Cluster Computing with Working Sets, HotCloud 2010.