

# Center for Research in Wireless Mobility and Networking (CReWMaN)

## Pervasively Secure Infrastructures (PSI) Grant No. IIS-0326505

NSF Fourth Annual Report (June 2006 - May 2007)

<http://crewman.uta.edu/psi>

### Abstract

The fourth year of the PSI project saw a major thrust in the area of Wireless Sensor Networks, Computer and Network Security, Pervasive Computing, Machine Learning, and Databases. Through our collaborative effort we have been able to propose a new security scheme in end-to-end data authentication with support for aggregation. Our approach is based on the concept of digital image watermarking and is superior to existing MAC based authentication schemes. We have also extended our research to include novel approaches for deadlock avoidance and routing for systems where multiple resources are available for the same task. Also, a new matrix formulation for decision-making in sensor networks using Dempster Shafer evidence theory has been developed. Considering the importance of ubiquitous healthcare applications, we are in the process of developing a framework that supports efficient context-aware data fusion for the building of distributed context-aware healthcare applications. Our framework provides a systematic approach to derive context fragments, and deal with context ambiguity in a probabilistic manner. The “Subdue” data learning and mining system has been further enhanced to learn concepts from relational data. This, we believe, will augment our ability in identifying potential threats in many security-related domains. Research in databases has resulted in efficient incremental search and ranking of complex patterns in text streams (iInfoSearch), Hierarchical reduction of graphs (HDB-Subdue) and identifying best subgraph using information theoretic models (e.g., MDL or minimum description length principle), frequent subgraph discovery (DB-FSG), mining of significant intervals and episode discovery, stream and event processing for monitoring, and use of event driven approach for information security. A newer version of “infoFilter” system incorporates most of this research.

## 1 Wireless Sensor Networks (WSNs) and Network Security.

*Faculty Involved:* Sajal Das, Frank Lewis, Kalyan Basu,

*Student Members:* H. Ammari, P. Ballal, H. J. Choe, P. De, J. Ho, S. Kundu, W. Zhang.

*WebPage:* <http://www.crewman.uta.edu/psi>, <http://arri.uta.edu/acs/>

### Dynamic Resource Allocation for Air/Ground Mobile Sensor Network with Localization

In unstructured environments, dynamic resource assignment is required for effective cooperation of robot teams. In some scenarios, robots are in charge of executing multiple missions simultaneously. This creates

risks of deadlock due to the presence of shared resources among various missions. The main contribution of this research is the development of a novel approach that combines the one step look-ahead deadlock avoidance policy with dynamic resource assignment. The dynamic resource assignment is achieved using greedy resource assignment for multi-mission robot teams in the framework of a matrix based discrete event controller.

## **Mission Programming for Deployable Wireless Sensor Networks**

In this work, we have applied a matrix-based discrete event controller (US Patent) to wireless sensor networks for mission programming, supervision in task sequencing, and resource assignment. The needs of WSN are not the same as other discrete event systems, and we modified the controller to allow for multiple missions, mission priority interrupts, and fast assignment of sensors without blocking phenomena. The WSN consists of some fixed unattended ground sensors, and some mobile sentry nodes that can vary their location to enhance the capabilities of the WSN in terms of repairing damage, compensating for faults, providing additional sensor information, and responding to detected events. The discrete event controller assigns the mobile nodes to tasks to carry out programmed missions, and has alarm capabilities when certain events are detected.

## **Collaborative exploration, Map Building, and Trust**

In this research, we explore the topics of path building with exploration, coordination of tasks and secure environment for robot team cooperation. Mobile wireless sensor networks are more vulnerable to security attacks than wired networks due to the broadcast nature of the transmission medium. Also, they have additional vulnerability as the mobile nodes are often placed in hostile environments where they cannot be physically protected. Since the base station is the gateway for the nodes to communicate with each other, compromising the base station can render the entire WSN useless. During the creation of WSN, each node is given a master key which is shared with the base station. All other keys are derived from this key. Mobile nodes can move in and out of the network, and hence have to be authenticated when they reenter the network. This gives rise to task planning where some mobile nodes have to intercept the incoming node and verify the master key. Such task planning can be easily done using a supervisory discrete event controller.

For example, the test bed at ARRI consists of mobile robots (Acroname Garcia) and a maze. There is a sentry robot which patrols the maze. This robot is localized using reference MIT Cricket ultrasound sensors. The sentry robot is programmed to perform obstacle avoidance, so that it can travel from the entry of the maze to the exit. In its journey, its path is recorded at the base station and sent back to the sentry robot. When another robot enters the grid (one which does not have obstacle avoidance), an event is triggered and the discrete event controller jumps into action. The guard robot intercepts the new mobile node, and pings for its network key. If the key is invalid, the new node is not allowed to enter the maze. If the key is valid, the sentry robot sends the path information to the new robot and allows it to pass through the maze.

This type of a network is an example of hybrid control, where the base station acts as the central supervisory controller and the robot navigation and obstacle avoidance form a lower level control. Toolkits in LabVIEW have been developed for easy task planning and programming of mobile nodes.

## **Aggregation Supportive Authentication in Wireless Sensor Networks: A Watermark Based Approach**

In wireless sensor network, security is a major concern for a plethora of applications, especially for those deployed in unattended or hostile environments. For secure data aggregation, current approaches have proposed enroute data authentication based on message authentication code (MAC), a keyed hash function. Usually sensor nodes append a MAC to an aggregate report as an endorsement. These MACs are checked by the intermediate nodes along the route to the sink. However, the high frequency of MAC checking along with complicated peer-to-peer key management schemes dramatically increase the overall system complexity. Moreover, the nature of MAC itself excludes any aggregation on the data as it will inevitably result in invalid MAC. This encourages to investigate solutions that support aggregation

supportive authentication in sensor networks. In order to overcome the limitations of MAC based authentication, we propose an end-to-end authentication approach with inherent support for aggregation. We visualize the sensory data gathered from the whole network at a certain time snapshot as an image. Under such visualization, every sensor node is viewed as a pixel with its sensory reading representing the pixels intensity. Thus, we can apply digital watermarking techniques to this “sensory data image. Specifically, the authentication information is modulated as watermark and superposed to the sensory data at the sensor nodes. The watermarked data can consequently be aggregated by the intermediate nodes without incurring any en-route checking. Upon reception of the sensory data the data sink is able to authenticate the data by validating the watermark, detecting whether the data has been altered and where it has occurred. Our simulation results show that the proposed scheme can perform aggregation supportive authentication. It can also detect different attacks that are launched in either spatial or frequency domain.

## Modeling Node Compromise Using Epidemic Theory

We have mathematically modeled the process of node compromise spread based on Epidemic Theory and studied the effects of various node deployments on the process. We assume the nodes in a sensor network to be securely communicating with each other using secret keys shared between them. In the event of a capture of a sensor node we assume that all its keys are known by the adversary. In such a situation, we observe the process by which a captured node gradually compromises other nodes by securely communicating with it and transmitting malware. Since a sensor network is generally static in nature, the assumption of homogeneous mixing of the individual nodes, which is done normally in a differential rate equation based formulation in Epidemic Theory, is not applicable. Therefore, we approach the problem from a random graph theoretic standpoint. The key parameters that we try to identify are the exact points when the node compromise process scales into an epidemic and affects the whole network. Moreover, since a random graph theoretic approach is unable to capture the temporal effects of the spread of this node compromise, we performed simulations to do that. In our simulation study, we observed the way the node compromise process behaves with time. We studied the process under two assumptions. One, when there is no node recovery process and second, when there is a recovery process underway.

## Coverage Issues in Wireless Sensor networks

We investigate coverage in the presence of obstacles. Obstacles are particularly realistic because of the presence of uneven surfaces, trees, hilly terrains inside the domain. We have seen that obstacles pose significant challenges, hence existing tools and techniques has to be extended considerably to solve coverage problems in presence of obstacles. In particular, we study the presence of obstacles in computing BCP(s, t) (Best Coverage Path between two points s and t) in a 2D field under surveillance by sensors. Consider a set of m line segment obstacles and n point sensors on the plane. For any path between s to t, p is the least protected point along the path such that the Euclidean distance between p and its closest sensor is maximum. This distance (the path’s cover value) is minimum for a BCP(s, t). We have developed two variants of the problem, the BCP(s, t) problem for opaque obstacles and the BCP(s, t) problem for transparent obstacles, based on variants of obstacle properties. We present two algorithmic results. For opaque obstacles, i.e., which obstruct paths and block sensing capabilities of sensors, computation of BCP(s, t) takes  $O((m^2n^2 + n^4)\log(mn + n^2))$  time and  $O((m^2n^2 + n^4))$  space. For transparent obstacles, i.e., which only obstruct paths, but allows sensing, computation of BCP(s, t) takes  $O(mn^2 + n^3)$  time and  $O(m^2 + n^2)$  space. Also, we provide mathematical proof of correctness for both of these results. We believe, this is one of the first efforts to study the presence of obstacles in coverage problems in sensor networks.

To solve these two problems, we use computational geometry based tools and techniques. In particular, to solve the BCP(s, t) problem for opaque obstacles, we have developed an algorithm that takes quartic-time, based on constructing a specialized dual of the Constrained and Weighted Voronoi Diagram. However for transparent obstacles, we have shown that computation of a BCP(s, t) is an easier problem and the hence can be done using visibility graph data structure. As an ongoing work, we are also investigating practical techniques such as approximation algorithms and heuristics for solving these problems efficiently. In addition, we are interested to consider an alternative problem which finds out

the set of sensors that can be reached in the plane given a source point  $s$  and a cover value  $c$  (techniques such as parametric search in computational geometry may be useful). We also plan to investigate other types of coverage problems in sensor networks in the presence of obstacles.

## Fault-tolerance measures of $k$ -covered wireless sensor networks

A fundamental aspect in the design of wireless sensor networks (WSNs) is their *functionality*. Because of their limited battery power (or *energy*), sensors may entirely deplete their energy or have their remaining energy below some threshold that will not help them function properly. These sensors are called *faulty sensors*. A WSN is said to be *functional* if there is a communication path between any pair of non-faulty sensors in the network. The notion of network functionality and thus fault tolerance strongly depend on the network connectivity. Precisely, a WSN is said to be *fault tolerant* if it remains functional in spite of the occurrence of sensor failures. Although network connectivity can be used to measure the fault tolerance of small-scale networks, it has a few shortcomings and hence is not appropriate for large-scale networks, such as  $k$ -covered wireless sensor networks, where every location in the field is simultaneously *covered* (or *sensed*) by at least  $k$  sensors (property known as  *$k$ -coverage*, where  $k$  is the sensing coverage). Indeed, traditional connectivity assumes that any subset of nodes can potentially fail at the same time and in particular all the neighbors of any node in the network. Thus, the fault tolerance of a network is upper-bounded by its minimum degree. We have investigated connectivity based on the degree of sensing coverage by studying  *$k$ -covered WSNs*. We have observed that to derive network connectivity of  $k$ -covered WSNs, it is necessary to compute the sensor spatial density required to guarantee  $k$ -coverage. More precisely, we have proposed to use a model, called Reuleaux Triangle, to characterize  $k$ -coverage with the help of the intersection of sensing disks of  $k$  sensors. We have proved that the minimum sensor spatial density required to guarantee  $k$ -coverage of a convex field is between  $2k/(\pi - \sqrt{3})r^2$  and  $2k/(\pi - \sqrt{3})r_{min}^2$ , where  $r$  and  $r_{min}$  are the radius and minimum radius of the sensing ranges of the sensors, respectively. We have also proved that network connectivity of  $k$ -covered WSNs is higher than their sensing coverage  $k$ . More precisely, we have proved that network connectivity  $\kappa$  of  $k$ -covered WSNs is between  $2\pi R^2 k/(\pi - \sqrt{3})r^2$  and  $2\pi R_{max}^2 k/(\pi - \sqrt{3})r_{min}^2$ , where  $r$  and  $r_{min}$  stand for the radius and minimum radius of the sensing ranges of the sensors, respectively, and  $R$  and  $R_{max}$  are the radius and maximum radius of the communication ranges of the sensors, respectively. This also implies that the network is fault-tolerant as it remains functional in spite of  $\kappa - 1$  sensor failures due to low battery power (or energy). This result is in sharp contrast with the previous work reporting that the connectivity of  $k$ -covered WSNs is equal to  $k$ . We have also proposed a new measure of fault tolerance for  $k$ -covered WSNs, called *conditional fault-tolerance*, based on the concepts of *conditional connectivity* and *forbidden faulty sensor set* that includes all the neighbors of a given sensor. To this end, we have proved that  $k$ -covered WSNs can sustain a large number of sensor failures provided that the faulty sensor set does not include a forbidden faulty sensor set. More specifically, we have proved that conditional network connectivity of  $k$ -covered WSNs is between  $4R(R+r)k/r^2$  and  $4R_{min}(R_{min}+r_{min})k/r_{min}^2$ , where  $r$  and  $r_{min}$  are the radius and minimum radius of the sensing ranges of the sensors, respectively, and  $R$  and  $R_{min}$  are the radius and minimum radius of the communication ranges of the sensors, respectively.

## References

- [1] V. Giordano, P. Ballal, F.L. Lewis, B. Turchiano, and J.B. Zhang, "Supervisory control of mobile sensor networks: math formulation, simulation, implementation", *IEEE Transactions on Systems, Man and Cybernetics*, Part B, Volume 36, Issue 4, Aug. 2006 Page(s):806-819.
- [2] A. Das, D. Popa, P. Ballal, and F. Lewis, "Data-logging and Supervisory Control in Wireless Sensor Networks", *ACIS Intl Journal of Wireless and Mobile Computing*, 2007, to appear.
- [3] P. Ballal, F. Lewis, J. Mireles Jr., and K. Sreenath, "Deadlock avoidance for free choice multi-reentrant flow lines: Critical siphons and critical subsystems", *Proc. Mediterranean Conf. Control & Automation*, Athens, Greece, June 2007.
- [4] P. Ballal, V. Giordano, P. Dang, S. Gorthi, and F. Lewis, "A LabView based test-bed with off-the-shelf components for research in mobile sensor networks", *Proc. ISIC*, Munich, Germany, October 2006.

- [5] P. Ballal, F. Lewis, J. Mireles Jr., and K. Sreenath, "Deadlock avoidance for free choice multi-reentrant flow lines: Critical siphons and critical subsystems", *Proc. Mediterranean Conf. Control & Automation*, Athens, Greece, June 2007.
- [6] W. Zhang, Y. Liu and S. K. Das, "Aggregation Supportive Authentication in Wireless Sensor Networks: A Watermark Based Approach", *IEEE WoWMoM*, 2007.
- [7] S. S. Ge and F.L. Lewis, "Autonomous Mobile Robots: Sensing, Control, Decision-Making, and Applications", *CRC Press*, 2006.
- [8] G. Vachtsevanos, F.L. Lewis, M. Roemer, A. Hess, B. Wu, "Intelligent Fault Diagnosis and Prognosis for Engineering Systems", *John Wiley, New York*, 2006, to appear.
- [9] S. Bogdan, F.L. Lewis, Z. Kovacic, and J. Mireles, "Manufacturing Systems Control Design: A Matrix Based Approach", *Springer-Verlag, London*, 2006, to appear.
- [10] H. M. Ammari and S. K. Das, "Coverage, Connectivity, and Fault Tolerance Measures of Wireless Sensor Networks", *Eighth International Symposium on Stabilization, Safety, and Security of Distributed Systems (SSS)*, pp. 35-49, Dallas, Texas, USA, November 17-19, 2006.
- [11] Habib M. Ammari and Sajal K. Das, "On Computing Conditional Fault-Tolerance Measures for k-covered Wireless Sensor Networks", *9th ACM/IEEE International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (IEEE/ACM MSWiM)*, Torremolinos, pp. 309-316, Malaga, Spain, October 2-6, 2006.
- [12] J. Wang, Y. Liu, and S. Das, "Asynchronous Sampling of Correlated Data in Wireless Sensor Networks", *IEEE WCNC 2007*, Hong Kong, China, Mar. 2007
- [13] S. Roy, G. Das, and S. Das, "Computing Best Coverage Path in the Presence Of Obstacles in a Sensor Field", *10th Workshop On Algorithm and Data Structures (WADS)*, accepted, 2007.

## 2 Pervasive Computing

*Faculty Involved:* Mohan Kumar, Behrooz Shirazi, and Sajal K. Das

*Student Members:* Nirmalya Roy, Mi Jeom Kim (Ph.D completed 2006), Kunal Shah, Nayantara Mallesh, Hitha Alex, Swaroop Kalasapur (Ph.D completed 2006), Kumarvel Senthivel and M H Ko (external student)

*WebPage:* <http://www.cse.uta.edu/pico@cse/>

### Middleware in Sensor Networks

In this project we introduce a Variable-Weighted Fair Queuing scheduling algorithm (V-WFQ) that provides network Quality of Service (QoS) by dynamically adapting to varying network traffic congestion at each router. Variable-Weighted Fair Queuing means that changes in congestion at a router will be reflected in a change in the relative priority among network flows. V-WFQ provides a prioritization scheme in which higher level Types of Service (ToS) flows dominate the network resources when network resources are constrained, leaving the lower level ToS flows with the remaining resources. This is accomplished through the altering of the flows relative priorities, using multiple forwarding queues. The industry standard Weighted Fair Queuing (WFQ) algorithm is a scheduler that uses a static priority mechanism with a predetermined number of forwarding queues. V-WQF, in addition to providing weighted fair queuing, provides the variability which allows the system to dynamically adapt to the current network load. We compare V-WFQ and WFQ using several QoS metrics including packet delay, packet loss, throughput, and weighted average system delay (WASD). Our results show that we are able to provide better QoS to higher level ToS flows when compared with WFQ.

### Resource Discovery in Ubiquitous Health Care

Ubiquitous computing is an emerging paradigm for health care environments, in which devices must blend into the background unobtrusively, collaborating to provide value-added services for users. Insufficient information flow and coordination are some of the main concerns in a ubiquitous health care environment. Resources are, therefore, essential to the success of this technology and, as a result, both resource

discovery and management play a vital role in sustainable ubiquitous deployment. In this project, we present a novel resource discovery mechanism for ubiquitous computing environments using Resource Index nodes (RINs) and mobile agents. The context information needed to process a query is identified and mobile agents are deployed to explore the ubiquitous environment for requested resources. We present a health care scenario and evaluate our proposed resource discovery method. Our proposed scheme has been developed for encompassing resource discovery on existing legacy systems and to facilitate information transfer between the ubiquitous health care system and the existing systems.

Ubiquitous computing is a challenge for the design of a middleware framework. Resource constraints, mobility, heterogeneity, scalability are just a few issues that have to be addressed. Such a middleware has to be tailored to the application scenario as well as the target platform. It is therefore effectual that the framework has to be built from minimal fine-grained components, and the system structure should also be highly configurable. A middleware developed for a ubiquitous computing environment has to be service-oriented and should be able to endure limited device capabilities. Fault-tolerance, integration, reconfiguration and usability are some of the innate characteristics of such a middleware. Another important feature of ubiquitous computing systems is their dynamic nature owing to the myriad number of devices deployed in a ubiquitous environment, along with the inherent mobility of these devices.

We aim at devising a scheme of data fusion which can facilitate expedited medical assistance from the point of request to the health care environment. At the same time, we strive to reduce privacy risks and aim at a coherent system which can anticipate such issues.

## Models for Service Composition

We have developed a service provisioning framework that is not only flexible in defining and generating services, but also adaptive to the environment. The service provisioning mechanism based on the community computing framework captures the essence of devices and services around users, and facilitates the creation and composition of flexible and adaptive services. An elegant mechanism for modeling the framework has been developed and several example scenarios investigated to demonstrate and validate the model.

The framework uses graph theoretic and sub-graph matching techniques for representing, creating, composing and operating services in pervasive environments. Further, we demonstrate the applicability of the proposed framework through the development of prototype models. Our investigations reveal that the proposed framework can be used to deploy middleware services swiftly and effectively in pervasive environments. We have developed graph based techniques to represent services and build composite services.

The main contributions in this task are: i) the development of Seamless Service Composition mechanism (SeSCo) for creating, composing and maintaining services in pervasive computing environments; ii) development of the LATCH protocol to create a hierarchical, distributed service maintenance mechanism in heterogeneous, dynamic environments; and iii) a methodology for evaluating service oriented architectures in pervasive environments.

Service provider communities for the management of communication networks (traditional as well as mobile) are under development to enhance TCP/IP and mobile IP performances. This research work assumes the existence of active elements in the communication network. Software agents and communities of agents execute on active elements to carry out specific goals/ missions. The scheme, called adaptive networking services has been prototyped for networks comprising heterogeneous devices and communication channels.

## Service Overlay Network for Pervasive Services

A lightweight service provisioning mechanism called, PerSON (Service Overlay Network for Pervasive Environments) has been developed to abstract the details of service provision and utilization in a pervasive environment. PerSON constructs an ad hoc service overlay network in the pervasive environment based on service requests. PerSON exploits the efficiency obtained by merging service discovery with route discovery in the overlay network where reactive routing protocols are used. Since PerSON stores only minimal information about discovered services and the neighboring devices, it is suitable for resource constrained devices. In particular, PerSON provides the overlay network for the community-computing concept introduced in Pervasive Information Community Organization (PICO). We also describe the

implementation of a prototype for emergency response system (ERS). The services and applications developed using PerSON may be hosted on devices with different computing capabilities. The devices may be connected to different physical or logical networks. A service overlay network is created by the framework based on service provision and utilization. The services in PerSON can be discovered and utilized by other services and applications, without dealing with the complexity of the underlying network. The PerSON framework is not dependent on a specific development language or operating system. The reference implementation of PerSON is developed using Java. The J2SE version of PerSON can be executed in powerful devices like desktops and laptops. The J2ME version can be executed in resource constrained devices like PDAs and cell phones. The J2SE version supports both TCP/IP and Bluetooth networks where as the J2ME version supports only Bluetooth networks. The reference implementation provides simple Java APIs to create, discover and utilize services.

The implementation of PerSON is used to provide the overlay network for the PICO middleware for pervasive computing. The main components of PICO are devices, and intelligent agents called delegents. PICO creates mission oriented dynamic and static and communities of delegents that perform tasks for the users and devices.

In PerSON, the directory of available services in the network is distributed. Each device stores only the information about the provided services. Other devices cache the information for the discovered services and the complete route to those services. Since there is no central repository, PerSON requires less memory to store the minimal service information, compared to the service-coordinator based approaches. PerSON uses simple text to describe services. The memory required to cache the discovered services is considerably reduced. A typical record in the service table requires 16 bytes for the service identifier, 8 bytes to specify the available time, 16 bytes the device identifier and say, 256 bytes for the service description. A total of 296 bytes is required to cache the information of a service. A typical record in the route table for a route with 3 hops requires 16 bytes for the destination device identifier, 1 byte for the length of the route and 32 bytes for the route. A total of 47 bytes is required to cache the route. A record in the device table that stores the information about another device on an IP v4 network requires 16 bytes for the device identifier, 8 bytes for the available time and 6 bytes for the IP address and the port number. A total of 30 bytes is required to cache the device information. The reference implementation of PerSON is capable of bridging IP networks and Bluetooth networks. Devices connected to only Bluetooth network can communicate with devices connected only to an IP network using any device that is connected to both networks. The bridging is done at the application layer above the TCP/IP stack and the Bluetooth stack. The device willing to route the messages should include the router component of PerSON stack. The router component simply forwards the message to the next hop in the route.

In order to support heterogeneous networks, the messages have to be routed in the overlay network. Dynamic source routing protocol is used in the overlay network for routing messages in the ad hoc environment. The route discovery is merged with service discovery. The route information is piggybacked in the service discovery query messages. A device receiving the query message knows the route to the device that requires the service. Similarly a device receiving the response message knows the route to the service provider. The service and route information are cached in each device. The complete route is specified in each message. The route includes only the list device identifiers of intermediate devices. Each intermediate device is responsible for connecting to the next hop in the route using the physical network connections.

A prototype for enhanced Emergency Response System (ERS) is implemented by employing services developed using the PICO middleware framework. Prototypes will be developed for secure environments and crisis management scenarios.

## **MidFusion: Middleware for Information Fusion**

Information acquired by large number of heterogeneous sensors needs to be integrated in a proactive, intelligent, and situation-aware manner to predict the occurrence of events (including security) in the PSI framework. In this project, we investigate the applicability of sensors by deploying collaborating software agents that meet the needs of dynamic applications. Two major challenges for proactive and real time collaboration among agents are (1) heterogeneity of sensors, information representation and granularity and (2) fusion of uncertain, redundant, complementary and time sensitive information from various sensors. We investigated the coupling of sensors and associated agents for real time information fusion and decision making in distributed and dynamic applications. The agents cooperate in real-

time to make intelligent and informed decisions using Bayesian Network reasoning. We have developed and demonstrated a learning based approach to effectively measure the confidence in cooperating agent observations. This work resulted in the development of an adaptive middleware architecture called MidFusion to facilitate information fusion in sensor network applications. MidFusion discovers and selects the best set of sensors or sensor agents on behalf of applications (transparently), depending on the quality of service (QoS) guarantees and the cost of information acquisition. The mechanism to select the best set of sensors using the principles of Bayesian and Decision theories has been developed. A sensor selection algorithm (SSA) for selecting the best set of sensors has been developed.

A large class of fusion problems can be decomposed into smaller fusion problems. Fusion is backward reasoning from symptoms (observable) to events (unobservable). Causal models that link symptoms to events facilitate clear problem decomposition. Complex fusion problems can be accomplished through hierarchies of simpler tasks each corresponding to partial estimation problem. Besides most information sources are typically very much hierarchical in nature. Problem decomposition means decomposition of the fusion problem into ordered sub tasks. Multi agent systems are suitable because they allow for encapsulation of sub tasks. Simple building blocks of the distributed and hierarchical information systems can be implemented through agents of different types dynamically organized into systems.

## Temporal Fusion using Dynamic Time Warping

Traditionally sensor fusion processes only concern fusing across raw data, features or decisions at specific points of time. However recently, there is a growing interest in inferring the behavioral aspects of environments or objects that are monitored by multi-sensor systems, rather than just their states at specific points in time. In order to infer environmental behaviors, it may be necessary to fuse data acquired from i) geographically distributed sensors at specific points of time and ii) specific sensors over a period of time. Fusing multi-sensor data over a period of time (also known as Temporal fusion) is a challenging task, since the data to be fused consists of complex sequences that are multidimensional, multimodal, interacting, and timevarying in nature. Additionally, performing temporal fusion efficiently in realtime is another challenge due to the large amounts of data to be fused. To solve this, we developed a robust and efficient framework that uses Dynamic Time Warping (DTW) as the core recognizer to perform online temporal fusion on either the raw data or the features. We evaluated the performance of the online temporal fusion system on two real world datasets: 1) accelerometer data acquired from performing two hand gestures and 2) a benchmark dataset acquired from carrying a mobile device and performing the predefined user scenarios. Performance results of the DTW based system are compared with those of a Hidden Markov Model (HMM) based system. The experimental results from both datasets demonstrate that the proposed system outperforms HMM based systems, and has the capability to perform online temporal fusion efficiently and accurately in realtime.

## Resource Management in Wireless Sensor Networks

In wireless sensor networks, resource-constrained nodes are expected to operate in unattended highly dynamic environments. Hence, the need for adaptive and autonomous resource/task management in wireless sensor networks is well recognized. We have developed Distributed Independent Reinforcement Learning (DIRL), a Q-learning based framework to enable autonomous self-learning/adaptive applications with inherent support for efficient resource/task management. The scheme based on DIRL, learns the utility of performing various tasks over time using mostly local information at nodes and uses the utility value along with application constraints for task management by optimizing global system-wide parameters like total energy usage, network lifetime etc. The feasibility of the proposed scheme has been demonstrated through an object tracking application. We carried out simulation studies to demonstrate the feasibility of our approach and compare its performance against other existing approaches. In general for applications requiring autonomous adaptation, we show that DIRL on average is about 90% more efficient than traditional resource management schemes like static scheduling without losing any significant accuracy/performance.

## Architecture for Deploying Services in Heterogeneous Pervasive Environments

We developed VSD (Service Discovery based on Volunteers), a service discovery architecture/protocol for heterogeneous and uncertain pervasive computing environments. The VSD architecture is centralized, but flexible. The servicing area of one directory can be overlapped with those of other directories. A small subset of the nodes called volunteers, perform the directory services willingly in the system. The volunteer refers is a duo comprising a software entity performing the directory services and a node hosting the software entity (a volunteer node). Relatively stable (less mobile) and capable (resourceful) nodes serve as volunteers. Cooperation among nodes is required for service discovery systems to perform seamlessly in pervasive computing environments. In adversarial environments, malicious nodes may harm honest ones. For instance, mean service providers may pretend to provide good services and make use of naive service requestors. To address this issue each nodes behavior is monitored to distinguish adversary nodes from good ones. Recognizing this need for secure operation, trust management has been incorporated into VSD architecture.

The VSD architecture is not limited to any specific routing protocols or physical network media. The volunteers may appear to have the same roles as directory agents (service proxies or brokers) in existing protocols. In our scheme, volunteer operations take place only on certain nodes volunteer nodes. Indeed, the volunteer mechanism has been introduced to exploit node heterogeneity and unevenness that prevails in most existing pervasive systems. The followings are unique features of VSD:

- Auto-configures the network with directory services without any administration or explicit leader election mechanism since any node (ideally stable and resourceful) can perform directory services.
- Provides reliability in uncertain environments through overlapped servicing areas of directories (clusters).
- Allows the establishment of trust relationships among directories, service requestors and providers in open environments without prior relationship or knowledge.

For secure interaction among participants in open networks, we present a hierarchical distributed trust management scheme tightly integrated with authentication protocols of the middleware architecture. We define trust notation and operators, and develop trust evolution processes. In the proposed trust management scheme, trust values of clients are maintained globally and consistently in their communities, resulting in the decrease in total overhead compared to the distributed approaches. Authentication protocols are employed to complete the trust management efficiently. The proposed security mechanisms operate transparently and perform autonomously through cooperation amongst nodes.

Simulation studies demonstrate that the proposed trust management protocols exhibit high efficiency and performance compared to existing distributed approaches. We validate that both belief and trust models achieve the basic goal of distinguishing good and bad nodes in malicious environments. The middleware architecture developed in the PICO project is exploited to incorporate existing trust and security mechanisms effectively. We have demonstrated the incorporation of security mechanism within service discovery as an application. However, the developed mechanism can be applied to any service or application that requires interactions among nodes.

### 2.1 Adaptive Network Service (ANS)

**The Architecture:** ANS is a logical community of collaborating software agents that reside and execute on a subset of network nodes. ANS community comprises ANS agents and ANS nodes. ANS agents monitor, collect and exchange information about network conditions among agents in the ANS community and ANS nodes provide services to incoming user requests. The information collected by each ANS agent is communicated with other ANS agents in the community. Exchanging information between agents allows ANS to get an overall view of network conditions. ANS uses this information to route user data flows across the network. Before start of service, ANS decides the best available ANS node to service the user request. ANS considers traffic conditions in different parts of the network while deciding the best available ANS node to service the user request. ANS also monitors resources utilization in different ANS nodes. ANS uses this information to route incoming user requests around congestion and towards resource high areas of the network. ANS tries to engineer traffic flow in the network to increase user performance at the same time maximize network resource utilization. ANS nodes provide requested

service to user flows and provision resources to user flows on request by request bases. ANS nodes offer functionality to alter, buffer and route each packet that flows through the node. ANS nodes register with the ANS agent in its vicinity. Each ANS agent monitors a number of nodes. When a user requests service from ANS, the contacted agent communicates with the agent monitoring the selected ANS node. The agent monitoring the ANS node communicates this request to the selected ANS node to notify the ANS node of the new user request. The information collected by an ANS agent is communicated with other ANS agents in the community. The information received from other ANS agents as well as information obtained from monitoring the local area is stored in a lookup table. Congestion indicators like mean TCP congestion window size allow ANS to know if the user flow encountered congestion in the network. By collecting such data from different locations, ANS can identify the least congested network area. ANS agents monitor congestion indicators while the data flow is in progress. Statistics like RTP receiver reports are collected from the receiver at the termination of RTP flows using ANS.

**Location:** ANS agents are distributed across the network and are strategically located to provide services and monitor network conditions. Agents are deployed on nodes on the network monitor network resources and current traffic conditions. ANS agents are present on a subset of network nodes. This is an advantage over other schemes like MPLS, DiffServ and IntServ wherein support from every router along the path from source to destination is a necessity. Moreover ANS does not require changes to existing router functionality.

**Service Requests:** A user wishing to use ANS contacts an ANS agent in its vicinity. The ANS agent consults the local lookup table and depending on the type of service requested and current conditions in the network, chooses the best ANS node to service the request. The decision is based on resource availability and network traffic conditions. The ANS agent communicates the address of the ANS node to the user. The user proceeds to use the ANS service by directly contacting the ANS node. **Resource Provisioning:** Many network services require resources to be reserved within the network before the start of data transfer. ANS provides resource reservation for user data flows using the ANS service. Resources are reserved before the first byte of data is sent and are freed once the data flow is complete. ANS nodes communicate resource utilization data to the ANS agent in its vicinity. ANS agents collect data of previous flows and compile information about performance and network conditions. This information can be used to route new incoming user requests. We are investigating the use of ANS for employing packet finger printing and providing authentication and anonymity.

### 2.1.1 Planned major tasks for 2007-2008

1. Develop a framework for fault-tolerant distributed computing in pervasive environments.
2. Develop a methodology for decision making in resource constrained sensor environments.
3. Develop context-aware adaptable version of Seamless Service Composition.
4. Incorporate authentication and privacy features into service composition.
5. Prototype development of pervasively secure environment.

## References

- [1] G. Pallapa, N. Roy, and S. K. Das, "Precision: Privacy Enhanced Context-Aware Information Fusion in Ubiquitous Healthcare", *First Workshop on Software Engineering of Pervasive Computing Applications, Systems and Environments (SEPCASE '07)*, in conjunction with ICSE 2007.
- [2] G. Pallapa and S. K. Das, "Resource Discovery in Ubiquitous Healthcare", *Second IEEE International Symposium on Pervasive Computing and Ad Hoc Communications (PCAC-07)*, May 2007.
- [3] Alex H, Kumar M and Shirazi B, MidFusion: An Adaptive Middleware for Information Fusion in Sensor Network Applications, *Information Fusion Journal, Special Issue on Information Fusion in Distributed Sensor Networks*, In Press.
- [4] Kim M, Kumar M, and Shirazi B, "Service Discovery using Volunteer Nodes in Heterogeneous Pervasive Computing Environments", *Elseviers Pervasive and Mobile Computing*, vol. 2, no. 3, pp. 313-343.

- [5] Ko M.H, West G, Venkatesh V, and Kumar M, "Online Temporal Fusion in Multisensor Systems using Dynamic Time Warping, *Information Fusion Journal*", In Press.
- [6] S. Kalasapur, M. Kumar, and B. Shirazi, "Dynamic Service Composition in Pervasive Computing Systems", *IEEE Transactions on Parallel and Distributed Systems*, In Press.
- [7] M J Kim, M Kumar, B A Shirazi and H J Chong, "A Novel Architecture for Provisioning Basic Services in Heterogeneous Pervasive Environments", *International Symposium on World of Wireless Mobile and Multimedia Networks (WoWMoM)*, Helsinki, June 18-21, 2007.
- [8] Kunal Shah and Mohan Kumar, "Distributed Independent Reinforcement Learning (DIRL) Approach to Resource Management in Wireless Sensor Networks", submitted to MASS 2007.
- [9] K. Senthivel, S. Kalasapur, and M. Kumar, "A Framework for Service Overlay Network in Pervasive Environments", submitted to EUC 2007.

### 3 Machine Learning, Data Mining and Computational Intelligence

*Faculty Involved* : Diane Cook, Lawrence Holder

*Student Members* : William Eberle (PhD, 2007), Janakiram Natarajan (MS, 2007), Amar Singh (MS, 2007), Nikhil Ketkar (PhD expected 2008), Yan Zhang (MS, expected 2008), Ashish Singh (MS, expected 2008), Saiparvathi Chinnychrishnan (MS, expected 2008)

*WebPage*: <http://www.subdue.org>

#### Mining Graph Data

The ability to learn concepts from relational data has become a crucial challenge in many security-related domains. For example, the U.S. House and Senate Intelligence Committees' report on their inquiry into the activities of the intelligence community before and after the September 11, 2001 terrorist attacks revealed the necessity for "connecting the dots", that is, focusing on the relationships between entities in the data, rather than merely on an entity's attributes. The ability to discover relationship-driven patterns can impact our ability to prevent future attacks and ensure national security. A natural representation for relational information is a graph, and the ability to discover previously-unknown patterns in such information could lead to a significant improvement in our ability to identify potential threats. This research investigated approaches to improve the scalability and effectiveness of the Subdue graph-based relational learning system.

In the past year we have continued to investigate several extensions to Subdue that will allow it to more efficiently and effectively identify patterns in graphs representing data from security-related domains. Specifically, we have developed a new approach for discovering anomalies in graphs and applied it to the task of detecting illicit cargo. Second, since every object of interest in security domains is typically related to many other objects, we have investigated methods for finding patterns in one large graph, where certain nodes are labeled with a class value, as opposed to having to extract subgraphs from the larger graph, possibly cutting important relations in the process. Finally, we have released a new version of Subdue that includes the ability to process graphs incrementally, as new data becomes available. Below, we first briefly describe the Subdue approach and then discuss these latest developments.

#### Substructure Discovery

Subdue accepts as input directed or undirected graphs with labeled vertices (nodes) and edges (links). As an unsupervised discovery algorithm, Subdue searches for a substructure, or subgraph of the input graph, that best compresses the input graph. Once the search terminates and Subdue returns the list of best substructures, the graph can be compressed using the best substructure. The Subdue algorithm can be invoked again on this compressed graph, generating a hierarchical description of discovered substructures.

To allow slight variations between instances of a discovered pattern, Subdue applies an inexact graph match between the substructure definition and potential instances. Subdue's run time is polynomial in the size of the input graph. Substructure discovery using Subdue has yielded expert-evaluated significant

results in domains including predictive toxicology, network intrusion detection, earthquake analysis, web structure mining, and protein data analysis.

Subdue can also be used to learn concepts that distinguish examples of different classes. New examples that contain the discovered substructures are classified as positive examples, otherwise they are classified as negative examples.

## Recent Activities

### Graph-based Anomaly Detection

Using information theoretic, probabilistic and maximum partial substructure approaches, we have developed three novel algorithms for analyzing graph substructures for the purpose of uncovering all three types of graph-based anomalies: modifications, insertions and deletions. The key to the algorithms lies in our definition of an anomaly. Basing our definition on the assumption that an anomaly is not random, for instance in the case of committing fraud, we believe that this type of anomaly should only be a minor deviation from the normal pattern. Because anyone who is attempting to commit fraud or hide devious activities would not want to be caught, it only makes sense that they would want their activities to look as real as possible. For example, the United Nations Office on Drugs and Crime states the first fundamental law of money laundering as "The more successful money-laundering apparatus is in imitating the patterns and behavior of legitimate transactions, the less the likelihood of it being exposed". Thus, if some set of data is represented as a graph, any nefarious activities should be identifiable by small modifications, insertions or deletions to the normative patterns within the graph.

Our first algorithm uses the minimum description length principle to determine the normative pattern, and from that pattern, find patterns that while structurally similar, have some relational deviation that is within an acceptable level of change. By determining what substructure minimizes the description length of the graph, we are able to calculate the cost of transformation for instances within the graph that do not exactly match the discovered normative pattern, and as such, are indicative of an unexpected change.

Our second algorithm again determines the normative pattern as the one that minimizes the description length of a graph, but instead of looking at changes to this pattern we examine the probability of extensions to the pattern. If the normative pattern does not completely compress the graph, meaning there are other vertices and edges connected to the normative pattern, we examine each of these extensions in terms of the probability of their existence. If the probability of existence is low enough, we mark the instance as anomalous. We can then compress the graph by this anomalous instance, and repeat the process until there are no more extensions to the anomalous substructure.

Our third algorithm uses a trail of pattern expansion to discover the instances that are structurally deficient from the normative pattern. When we attempt to discover the pattern that minimizes the description length of the graph, we maintain a parental relationship between the structures. Once we have discovered the normative pattern, we traverse these relationships to find the instance that is the maximum partial substructure. In this case, we are looking for patterns that are unable to extend to the normative pattern, and are a maximal representation of that normative pattern. In other words, the maximum partial substructure is found in the instance that requires the fewest additions (if they would have existed) for transforming the instance into an instance consisting of the normative structure.

We validate all three approaches using synthetic data, verifying that each of the algorithms on graphs and anomalies of varying sizes, are able to detect the anomalies with very high detection rates and minimal false positives. We then further validate the algorithms using real-world cargo data and actual fraud scenarios injected into the data set with 100 algorithms demonstrates the usefulness of examining a graph-based representation of data for the purposes of detecting fraud.

### Learning from Supervised Graphs

A supervised graph is one large graph in which subgraphs are labeled as being positive or negative examples of some phenomenon. Our goal is to find patterns that can distinguish positive from negative examples (supervised learning) without having to extract examples from the large graph, and thus avoiding the possibility of cutting important relationships. Consider an example from a money laundering domain which comprises data about individuals, institutions and the transfer of funds among them. This data is multi-relational in nature and a graph-based representation provides a natural way to model this

domain. Here, vertices represent individuals and institutions while edges represent the transfer of funds among individuals and institutions.

In such a graph assume that we know certain individuals, institutions and the transactions between them to be fraudulent in nature. We also know certain individuals, institutions and the transactions between them to be innocent in nature. These individuals and institutions along with the transfer of funds among them can be viewed as subgraphs in the graph representation of the multi-relational data from the money laundering domain. What is peculiar about these subgraphs is that they can be viewed to either possess or lack the property of being fraudulent in nature. We refer to these subgraphs as sites. Furthermore, the subgraphs known to possess some distinct properties are referred to as positive sites and the subgraphs known to lack some distinct properties are referred to as negative sites.

Given this graph containing positive and negative sites, we would like to identify the characteristics that distinguish the positive sites from the negative sites. The characteristics which distinguish positive sites from negative sites might be certain graph patterns in the proximity of sites. In our example it is likely that the transfer of funds among fraudulent individuals and institutions have a pattern different from the transfer of funds between innocent individuals and institutions.

This technique could also be used to identify previously unknown terrorist cells or threat groups using a communication graph and known terrorist cells. Another example the identification of previously unknown functional modules from gene interaction graphs and known functional modules. Among the current approaches for mining graph-based data, only the inductive logic programming approach can be applied to this task to a certain extent. ILP systems can learn predicate definitions analogous to a site in the case where all sites are isomorphic to each other. Our experiments show that our approach performs better than ILP while learning large concepts and while learning from large graphs. In the case where the sites consist of nonisomorphic subgraphs with varying number of vertices and edges, ILP performs poorly as this involves learning a predicate definition containing a list which can have a variable length, and is computationally expensive.

In this work, we characterize this task as a supervised learning problem and present a heuristic algorithm for the same. Experimental comparison with the ILP system CProgol on real world and artificial datasets evaluates the ability of the approach in uncovering interesting patterns. We are evaluating this approach in a number of real world and artificially generated datasets.

## Incremental Graph Discovery

A challenge that arises in applying Subdue to security data is processing structural data that arrives in incremental blocks. We had previously developed a method that avoids reprocessing the whole accumulated graph, but maintains summary statistics for each increment and processes the new increments independently. This actually results in overall runtime improvement for Subdue and allows near-real-time handling of streaming structural data.

In this past year, we have released a new version of Subdue (5.2.1) that adds the incremental processing algorithm into the main Subdue algorithm. Graph increments are provided as a sequence of graph files, and the incremental processing is invoked using an additional command-line option.

## References

- [1] W. Eberle and L. Holder, “Anomaly Detection in Data Represented as Graphs”, *Intelligent Data Analysis*, 2007 (to appear).
- [2] J. Kukluk, L. Holder, and D. Cook, “Inference of Node Replacement Graph Grammars”, *Intelligent Data Analysis*, vol. 11, no. 4, 2007.
- [3] D. Cook and L. Holder (editors), “Mining Graph Data”, *John Wiley and Sons*, 2006.
- [4] W. Eberle and L. Holder, “Mining for Structural Anomalies in Graph-Based Data”, *International Conference on Data Mining (DMIN-07)*, 2007.
- [5] J. Kukluk, L. Holder, and D. Cook, “Inference of Edge Replacement Graph Grammars”, *Proceedings of the Twentieth International Conference of the Florida AI Research Society (FLAIRS)*, 2007.
- [6] N. Ketkar, L. Holder, D. Cook, R. Shah, and J. Coble, “Mining in the Proximity of Subgraphs”, *Proceedings of the ACM KDD Workshop on Link Analysis*, 2006.

- [7] A. Singh and L. Holder, "Classification of Threats via a Multi-sensor Security Portal", *Proceedings of the IEEE Intelligence and Security Informatics Conference*, 2006.
- [8] W. Eberle and L. Holder, "Detecting Anomalies in Cargo Shipments Using Graph Properties", *Proceedings of the IEEE Intelligence and Security Informatics Conference*, 2006.

## 4 Mobile Database

*Faculty Involved* : Ali Hurson

*Student Members* : Mark Wenstrom (completed Ph.D comprehensive examination), Bo Yang (completed Ph.D, 2006), Xing Gao (completed MS, 2006), M. Ontang, Evans Jean, H. Hsu (completed Ph.D comprehensive examination), A. Tangpong.

*WebPage*: <http://www.cse.psu.edu/gis/>

### On demand based Services

Typically, a multidatabase system consists of a global component and a collection of local components. The global component hides the underlying heterogeneity and provides users a uniform global information access method. In order to preserve local data autonomy while providing full database functionality, the global component usually maintains a global schema that contains local schema information. Problems arise as the size of the multidatabase and the global schema grows. Maintaining and manipulating multiple copies of large global schema in a distributed environment is problematic.

Within the scope of the multidatabases, we proposed an elegant solution (The Summary Schemas Model SSM) for large-scale organization that addresses the problems associated with global schema approaches. The SSM is designed to support the identification of semantically similar/dissimilar data entities. The model maintains a hierarchical meta-data based on the semantics of the access terms exported from underlying local databases. This meta-data is used to intelligently resolve imprecise as well as precise queries.

We have simulated and prototyped SSM and evaluated its advantages as follows:

- SSMs meta-data is by orders of magnitude smaller than the meta-data generated by the Global-schema approach
- SSM preserves local autonomy
- SSM provides good system scalability
- SSM offers a good performance (search time)
- SSM allows resolution of imprecise queries.

However, the client/server prototyped model showed:

- Lack of portability,
- Lack of stability,
- Relying on network connectivity.

The aforementioned shortcomings and the trend of the advances in technology motivated us to expand the scope of the SSM in the following directions:

- Adding wireless communication and mobility to the SSM infrastructure
- Application of the Mobile agents as a means to carry query processing
- Adaptability of the SSM to support multimedia data
- Allowing each local database to be represented as a community of data sources.

## **Agent-based data retrieval for Mobile Multidatabase**

After witnessing the success of many mobile agent applications, we proposed and implemented a new infrastructure MAMDAS Mobile Agents within the framework of Mobile Data Access System. MAMDAS combines the merits of SSM and mobile agents in building a distributed large-scale information access systems. It aims to achieve higher performance, while providing special support for mobile users. Our experimental results have shown that MAMDAS is 6 times faster than the client-server based SSM prototype, because of the reduced network traffic. Moreover, MAMDAS demonstrated great scalability, portability, and robustness.

## **Agent-based Transaction Management for Mobile Multidatabase**

Transaction management is not a trivial task in distributed and multidatabase environment. The issue becomes more problematic when mobility and wireless communication are added to the infrastructure due to the technological constraints. Agent technology can be used to remedy the technological limitations in managing the transactions. We proposed an Agent-based Transaction Management for Mobile Multidatabase (AT3M) system. AT3M applies static and mobile agents to manage the transaction processing in mobile multidatabase system. It enables a fully distributed transaction management, accommodates mobility of the mobile clients and mobile data sources, and allows global subtransactions to process in parallel. The proposed Agent-based Transaction Management (AT3M) system is under development and analysis.

## **Power management**

We developed an adaptive application-driven power management (AADPM) strategy with online idle period length distribution learning capability for the IEEE 802.11b WLAN. We evaluated its performance in comparison with other power management strategies using the network simulator NS2. We simulated both the single user and multiple user scenarios. Experimental results have shown that, compared with other power management methods, AADPM achieves the highest energy saving in all cases and it demonstrates strong adaptability to network congestion.

## **Location dependent data processing**

The scope of our infrastructure was extended to accommodate location dependent data processing/services, location aware data processing/services, and continuous query processing in a mobile environment with an eye toward the improvement of performance metrics such as access time, power consumption, etc. Two types of the queries, namely, window query and nearest neighbor query have been considered. Algorithms have been developed and simulated to improve performance by estimating the validity region, semantic caching, and proxy caching.

## **Image Retrieval in ad hoc network**

Mobile ad hoc networks have gained more and more research attentions by provisions of wireless communications without location limitations and pre-built fixed infrastructure. Because of the absence of any static support structure, ad hoc networks are prone to several limitations such as bandwidth, connectivity, and power. Multimedia retrieval is a challenging task in wireless ad hoc networks because of the multiple limitations. We investigated data content distribution to facilitate content-based multimedia retrieval in ad hoc networks. Motivated by this data organization methodology, we proposed a logic-based content summary framework that is able to represent semantic contents of multimedia data using concise logic terms. Furthermore, we built a virtual infrastructure to cluster mobile nodes according to their semantic contents. The proposed framework was simulated and analyzed based on various performance metrics.

## **Broadcast based services**

Many applications are directed towards public information. The reduced bandwidth of the wireless environment places limitations on the rate and amount of communication. Broadcasting is a potential solution to this limitation. The main advantage of broadcasting is due to the fact that it scales up as the number of users increases, eliminating the need to multiplex the bandwidth among users accessing the

air channel. In addition, broadcast channel can be considered as an additional storage available over the air for the mobile clients. Finally, it is shown that pulling information from the air channel consumes less power than pushing information to the air channel. Broadcasting is an attractive solution, because of the limited storage, processing capability, and power sources of the mobile unit. Within the scope of broadcasting one needs to address three issues:

- Broadcast contents
- Network latency
- Power consumption of the mobile unit

Application of indexing and broadcasting over parallel air channels are solutions that have advanced in the literature in order to reduce power consumption and access latency. However, broadcasting over parallel air channel brings the issue of conflict. Conflict occurs when two requested data objects can not be accessed during the same broadcast cycle. Increase power consumption and increase access latency are the natural by product of conflicts. To minimize the effect of conflicts on both access latency and power consumption, we developed access scheduling order that minimizes the number of passes over the parallel channels. We investigated the application of heuristics to schedule retrieval of data items from the parallel air channels. However, by the very nature of the heuristic rules, in some instances, the algorithm does not generate an efficient access order and/or reduced power consumption. We further extended the scope of our heuristic approaches and developed several new scheduling algorithms that can find the minimum number of passes and channel switches. The proposed scheduling algorithms have been simulated and their validity have been demonstrated based on several performance metrics such as the power consumption, access latency, and the number of passes over the parallel air channels.

### **On-going Activity and Planned Work for year 2007-2008:**

- **Security of Mobile agent:** Despite the fact that mobile agents have received increasing attention in various research efforts, the use of the paradigm in practical applications has yet to fully emerge. With the presence of infrastructure to support the development of mobile agent applications, security concerns act as the primary deterrent against such trends. Numerous studies have been conducted to address the security issues of mobile agents with a strong focus on the theoretical aspect of the problem. We will attempt to bridge the gap from theory to practice by analyzing the security mechanisms available in Aglet. We herein propose several mechanisms, stemming from theoretical advancements, intended to protect both agents and hosts in order to foster the development of business applications that fully exploit the benefits of agent technology. The proposed mechanisms lay the foundation for implementation of application specific protocols dotted with access control, secured communication and ability to detect tampering of agent data. We demonstrate our contribution through application scenarios of a prototyped Information Retrieval system.
- **MAMDAS and Pervasive Computing:** The use of MAMDAS can be extended to the pervasive computing environment, where computers and databases are pushed to the background and services are provided to users without being specifically requested. Smart classrooms that can automatically load lecture slides for professors according to the course schedule and syllabus exemplifies the idea of pervasive computing. We envision that the SSM can be used as the backbone knowledge base and autonomous mobile agents can act as user representatives who actually perform tasks on users behalf. Several difficulties must be addressed in this environment: i) seamlessly integrate heterogeneous networks, ii) implant human intelligence in agents, and iii) introduce user incentives.
- **MAMDAS and Sensor network:** The fast growth of sensor networks has attracted lots of research attention. Several prominent challenges in sensor networks include i) sensors have extremely scarce resources and short lifetime, ii) sensors have almost no physical protection, and iii) sensor network topology changes frequently. We believe the execution autonomy and decision making capability of mobile intelligent agents can alleviate the aforementioned problems.
- **Power-Aware Scheduling of Data Retrieval from Indexed Parallel Broadcast Channels:** To further improve power consumption and access latency, the scope of our research will be extended to include duplication of popular data items. The proposed scheduling algorithms with and without replicated data items will be simulated and simulation results will be presented and analyzed. It is

expected that duplication of data items has further impact on the power consumption and access latency.

- **A security framework for ad hoc networks:** We are intended to study and address the issues of network access control, attack source traceback, and network node integrity measurement, within the scope of the ad hoc networks. Several protocols are intended to develop and analyze.

## 4.1 Education and Outreach

Based on the aforementioned activities:

- A tutorial entitled, Application of Mobile agent in global information sharing at University of Pisa, Italy.
- Some of the research results presented in a talk entitled, Conflict resolution in Parallel Broadcast Channels at University of North Texas.
- Some of the research results presented in a talk entitled, Application of Mobile Agents in Information Retrieval Systems at University of Idaho.
- Some of the research results presented in a talk entitled, Power management in Mobile devices at University of Arkansas.
- Some of the research results presented in a talk entitled, Data dissemination through broadcast channels at Kansas State University.
- Research results presented in various ACM/IEEE/International conferences.

## References

- [1] Jiao Y., Hurson A.R., and Shirazi B., "Adaptive Application-Driven WLAN Power Management", *Journal of Pervasive and Mobile Computing*, vol. 3, no. 3, pp. 255-275, 2007.
- [2] Hsu H.-Y., Zhu S., and Hurson A.R., "LIP: A Lightweight Inter-layer Protocol for Preventing Packet Injection Attacks in Mobile Ad-Hoc Network", *International Journal of Networks and Security*, Special Issue on Cryptography in Networks, 2006.
- [3] Jiao Y. and Hurson A.R., "Energy-Efficient Wireless Information Retrieval", *Journal of Computer and System Sciences*, Special issue on Network-Based Computing, 2007.
- [4] Jean E., Jiao Y., Hurson A.R., and Potok T.E., "SAS: A Secure Aglet Server", *Computer Security Conference*, April 11-13, 2007.
- [5] Hsu Hung-Yuan and Hurson A.R., "PEAN: A Probabilistic Energy Aware Neighbor Monitoring Protocol for Mobile Ad Hoc Networks", *IEEE International Symposium on Pervasive Computing and Ad Hoc Communications, (PCAP)*, 2007.
- [6] Hsu H.-Y., and Hurson A.R., "LIP: A Lightweight Network Access Restriction Protocol for Preventing Packet Injection Attacks in Mobile Ad-Hoc Network", *International Symposium on Systems and Information Security*, 2006.
- [7] Reed J.W., Jiao Y., Potok T.E., Klump B.A., Elmore M.T., and Hurson A.R., "TF-ICF: A New Term Weighting Scheme for Clustering Dynamic Data Streams", *International Conference on Machine Learning and Applications, (ICMLA)*, pp. 258-263, 2006.
- [8] Yang B., Hurson A.R., Jiao Y., and Potok T.E., "Multimedia Correlation Analysis in Unstructured Peer-to-Peer Networks", *IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks*, pp. 332-338, 2006.
- [9] Jiao, Y., Hurson, A.R., and Potok, T., "Mobile Agent-Based Information Systems and Security", *Encyclopedia of Information Science and Technology*, 2nd edition, 2006.
- [10] Jiao, Y., Hurson, A.R., Potok, T.E., and Beckerman, B.G., "Integrating Mobile-Based Systems with Health Care Databases", *Web Mobile-Based Applications for Healthcare Management*, 2006.
- [11] Yang, B. and Hurson, A.R., "Location-Aware Caching for Spatial Queries in Dynamic Environments", *IEEE Wireless Communications and Networking Conference, (WCNC)*, 2006.

- [12] Ayyagari, P., Mitra, P., and Hurson A.R., “Efficient Object Retrieval from Parallel Air Channels in the Presence of Replicated Objects”, *International Conference on Mobile Data Management, (MDM)*, 2006.
- [13] Yang, B. and Hurson, A.R., “Multimedia Semantics Integration Using Linguistic Model”, *International Conference on Knowledge Discovery and Data Mining (PAKDD)*, pp 679-688, 2006.
- [14] Yang, B. and Hurson, A.R., Content-Initiated Organization of Mobile Image Repositories“, *IEEE/IFIP WONS2006*, 2006.
- [15] Sustersic, J., Hurson A.R., and Nickel, R. M. “An Analysis of Internet Data Update Behaviors”, *IEEE AINA2006 International Conference*, pp. 773-778, 2006.

## 5 InfoFilter: A Content-Based Information Filtering System

*Faculty Involved* : Sharma Chakravarthy

*Student Members*: R. Adaikkalavan, B. Kendai, A. Telang, C. H. H. Subramanian

*WebPage*: <http://itlab.uta.edu/>

### 5.1 Overview

As part of the research and development activity for PSI for the year ending in May 2007, a number of issues related to information security have been addressed. These include: efficient incremental search and ranking of complex patterns in text streams (iInfoSearch), Hierarchical reduction of graphs (HDB-Subdue) and identifying best subgraph using information theoretic models (e.g., MDL or minimum description length principle), frequent subgraph discovery (DB-FSG), mining of significant intervals and episode discovery, stream and event processing for monitoring, and use of event driven approach for information security. A number of MS (4) and PhD (1) theses as well as archival publications (2 journal papers and 13 refereed conference/workshop papers) have resulted from this project activity during this year. Below, we highlight the results of a few of the above activities.

### 5.2 Incremental InfoSearch (iInfoSearch)

Applying appropriate searching mechanisms to retrieve only relevant information becomes critical to avoid information overload (or retrieving very large number or portions of documents). *Information Retrieval* is the process of extracting relevant or useful portions of documents from a relatively static collection of documents based on a stream of incoming user patterns (or queries). In information retrieval, expressiveness of pattern (or query) specification by a user and its detection (or matching) play a significant role. In other words, in order to extract useful or meaningful information, users need to have the ability to specify complex patterns. A critical limitation of current search engines is that they can support only simple patterns or a simplistic combination of patterns using Boolean operators. Thus, current query languages are quite restrictive in their expressive power and need to be extended and generalized to address the specification of meaningful complex user patterns. On the other hand, ability to specify complex patterns will not be meaningful or effective without a correct and efficient mechanism for their detection in real-time. In this report we summarize, InfoSearch, a novel approach for expressive pattern matching over stored data. It allows users to specify complex patterns and matches these patterns over stored data. Complex patterns such as combinations of sequential, structural patterns, wild cards, word frequencies, proximity, Boolean operators and synonyms are formulated using the expressive pattern specification language, PSL.

### 5.3 Search for k and Top-k Patterns

Building upon our earlier work on InfoSearch [1, 2], we have extended it to identify K patterns where k is specified by the user. Many a times, it is not necessary to search for *all* patterns as it takes excessive amount of resources and time. The earlier algorithm to detect all patterns is likely to be limited by due to memory constraints as well as the size of the data in which the patterns are being identified.

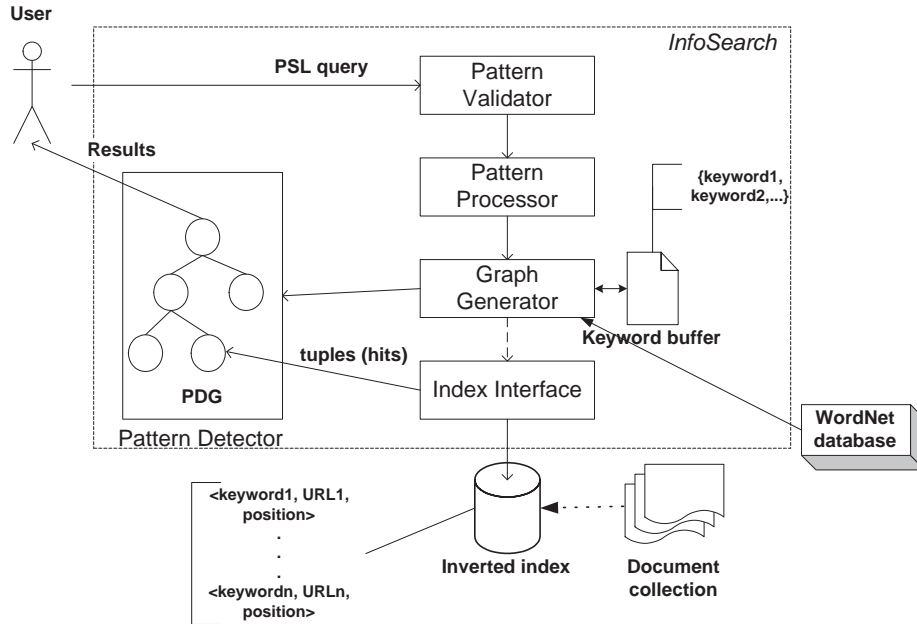


Figure 1: System Architecture

We have developed an approach and an algorithm to detect  $k$  patterns incrementally and stop the detection process as soon the required number of patterns are detected. Unlike the previous approach, data is retrieved from the index in small chunks in order to reduce the memory requirement for complex expression detection. It is sweeping, incremental algorithm and uses heuristics to determine how many instances of the same simple pattern occurrences needs to be sent in each sweep.

The system architecture shown in Figure 1 is the same used by earlier approaches but the algorithm is different. Currently, extensive experimentation is being conducted to compare all the three approaches [3].

In addition to merely finding first  $k$  patterns, it would be useful to identify top- $K$  patterns where the relevance of the pattern is well-defined. The need for top- $K$  search has been well established in the literature in text retrieval and currently in database retrieval. We have come up with a heuristics to attribute the utility of a pattern based on its tightness and have used it for the purpose of ordering the patterns identified. For example, if one is looking for a pattern "Iran" FOLLOWED BY/30 "bomb" (which asks for detecting a pattern which contains the word "bomb" with in distance of 30 words of "Iraq") and if there are 2 patterns – one with a separation of 25 words and another with a separation of 10 words, we rank the one with 10 word difference as higher. This is based on the intuition that one is looking for patterns that are closer to each other than apart. This idea has been extended to all operators and arbitrary expressions. The sweeping incremental algorithm has been modified to take ranking into consideration when detecting patterns.

## 5.4 Efficient and Scalable Main Memory Algorithms for Significant Interval and Episode Discovery

Growing interest in the field of data mining over the past few decades has resulted in a number of algorithms for processing various kinds of data including time-series data. In general, time-series data can be defined as an ordered sequence of values of a variable at aperiodic time intervals. Time-series data analysis is used in a variety of data-centric applications such as economic forecasting, sales forecasting, budgetary analysis, stock market analysis, inventory studies, census analysis and so forth. Time-series data is known for its huge volume. Hence, discovering useful nuggets of information from them is a challenging task.

A considerable amount of work [4] has been done to process and mine through the large collections of time-series data sets via sequential mining. Most of the existing and traditional sequential mining techniques [5, 6] use data in its original form as time points.

On the other hand, in real-world applications (e.g., predicting automating devices in a smart home), we are interested in identifying intervals with a high degree of certainty in which events happen, instead of specific time points. For example, it is useful to extract intervals of high activity from telephone logs to understand the network usage.

Similarly, for smart home [7] class of applications, it is useful to consider periods of high activity of the devices rather than their actual usage at various points in time, to infer the usage patterns of each device as well as interactions between different devices. Another characteristic of these traditional sequential mining algorithms is that they process the entire data set several times. For time-series data, the sheer size of the data set makes running an algorithm that makes several passes on the entire data set time consuming. The efficiency of these sequential mining algorithms can be improved by reducing the data set without changing the outcome. The above characteristics, as well as the work in [8] makes two things very clear.

1. for many applications, it is important to interpret the occurrence of events in intervals rather than at time points,
2. the efficiency of these sequential mining algorithms can be improved by reducing the data, and
3. reducing the number of passes and intermediate results generated will further improve the efficiency of these algorithms

The algorithms developed as part of this project can be extended to other domains which generates time-series data. The motivation for this work came from the MavHome (Managing An Intelligent and Versatile Home) project which is a multi-disciplinary research project at the University of Texas at Arlington (UTA) focused on the creation of an intelligent and versatile home environment [7]. We assume a time-series data set (generated by some application). The aim is to automate various activities in this environment such as when to turn ON a light, when to turn OFF the TV/Audio system, etc., to maximize the inhabitant comfort (or power management, security monitoring, etc.). This requires prediction of the device usage patterns using the usage data.

## 5.5 Our Focus

Given the task of discovering significant intervals and frequent patterns/episodes<sup>1</sup>, we divide our task into three phases:

1. identifying the best significant intervals (tightest intervals) based on characteristics provided by the user,
2. discovering interactions between events, to identify frequently occurring patterns/episodes of different sizes and strengths,
3. validating the frequent episodes/patterns detected (if need be)

This work contributions in each of the above three phases. Each of these phases have work done by Ambika [9], Sunit [10] and Dhawal [11] for their respective thesis. The aim of this thesis is i) to provide an efficient main memory algorithm for significant interval discovery, ii) to provide an efficient main memory algorithm for frequent episode discovery, iii) to redesign the frequent episode validation phase, and iv) to extend and improvise the performance of the previous works. A summary of the contributions of this thesis is shown in the Figure 2.

The contributions of this work [12, 13, 14, 15, 16] in each phases is highlighted in Figure 2 as brown shaded boxes. Each contribution is explained in detail in [12].

In this work, we have developed a OnePass algorithm to discover significant intervals in a single pass only. We have also developed a OnePass-AllSI algorithm to maximize the effectiveness of OnePass algorithm. Experimental results have shown that our algorithm outperforms other similar works [10, 17]. We have verified that the patterns discovered by Hybrid-Apriori algorithm [18] and episode discovery algorithm [7] are the same.

---

<sup>1</sup>Please note that patterns and episodes are used interchangeably

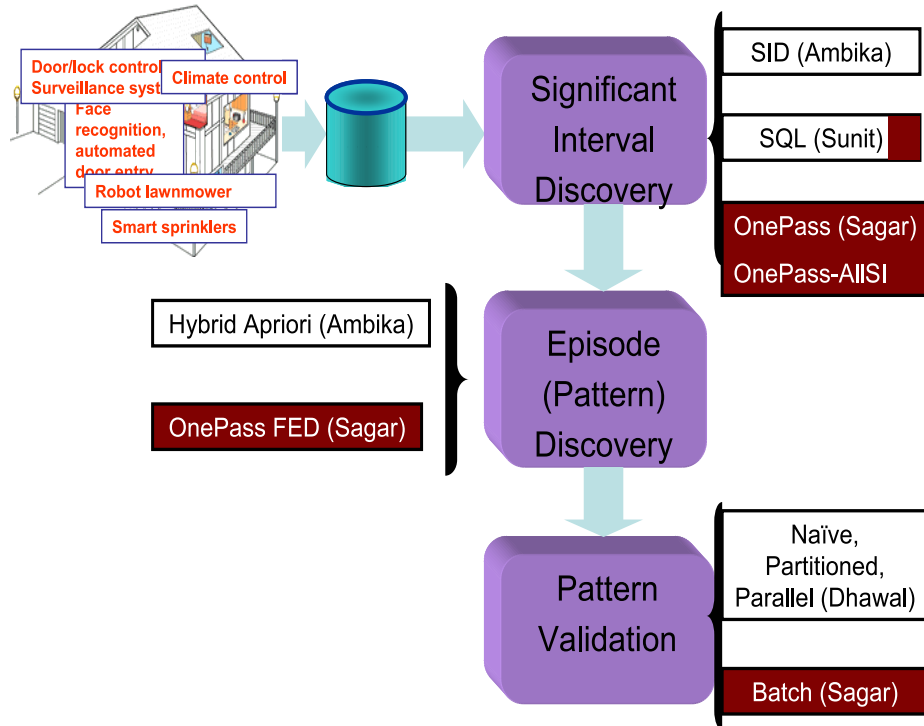


Figure 2: Different Phases of solution to smart home problem

We have also developed a OnePass FED algorithm to discover frequent episodes in a single pass in a much efficient manner. Experimental results have shown that, our algorithm significantly outperforms Hybrid Apriori algorithm (a frequent episode discovery algorithm) in terms of the time taken to find the frequent episodes.

## 5.6 Summary and Future Work

In this report, we discussed extensions to InfoSearch to detect K patterns and further how to rank patterns based on the proximity of simple pattern occurrences. We also briefly discussed our approach to significant interval discovery and episode discovery using main memory algorithms and compared their complexity and efficiency to earlier algorithms.

In addition to the work discussed above, we have worked on: best subgraph identification using minimum description length (HDB-Subdue) [19, 20], frequent subgraph discovery (DB-FSG) [21], use of flexible events for role-based access control (RBAC) [22, 23, 24], Selectively Monitoring customized content of web (HTML/XML) pages [25, 26] and integrating stream and event processing in a synergistic manner [27, 28, 29, 30, 31, 32, 33, 34], and information integration [35]

## 5.7 Proposed research for 2007-2008

We plan on continuing the above work for next year on all aspects of information security. Specifically, we are integrating stream and event processing by extending the notion of windows with semantic windows, understanding the relevance and applicability of scheduling and load shedding for event detection.

## References

- [1] L. Elkhalfa, “Infofilter: Complex pattern specification and detection over text streams,” Master’s thesis, The University of Texas at Arlington, 2004. [Online]. Available:

<http://itlab.uta.edu/ITLABWEB/Students/sharma/theses/Laali.pdf>

- [2] N. Deshpande, "Infosearch: A system for searching and retrieving documents using complex queries," Master's thesis, The University of Texas at Arlington, 2005. [Online]. Available: <http://itlab.uta.edu/ITLABWEB/Students/sharma/theses/Des05MS.pdf>
- [3] J. Thathireddy, "iInfoSearch: Incremental Detection and Ranking of Complex Text Patterns," Master's thesis, The University of Texas at Arlington, August 2007.
- [4] J. F. Roddick and M. Spiliopoulou, "A survey of temporal knowledge discovery paradigms and methods." *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, vol. 14, no. 4, pp. 750–767, Jul/Aug. 2002.
- [5] H. Mannila, H. Toivonen, and A. I. Verkamo, "Discovering frequent episodes in sequences." in *KDD*, 1995, pp. 210–215.
- [6] R. Srikant and R. Agrawal, "Mining sequential patterns: Generalizations and performance improvements." in *EDBT*, 1996, pp. 3–17.
- [7] D. J. Cook, G. M. Youngblood, *et al.*, "Mavhome: An agent-based smart home." in *PerCom*, 2003, pp. 521–524.
- [8] M. H. Böhlen, R. Busatto, and C. S. Jensen, "Point-versus interval-based temporal data models." in *ICDE*, 1998, pp. 192–200.
- [9] A. Srinivasan, "Significant interval and episode discovery in time-series data," Master's thesis, The University of Texas at Arlington, 2004. [Online]. Available: <http://www.cse.uta.edu/research/publications/Downloads/CSE-2003-39.pdf>
- [10] S. Shrestha, "Sql-based approach to significant interval discovery in time-series data," Master's thesis, The University of Texas at Arlington, 2005. [Online]. Available: <http://itlab.uta.edu/ITLABWEB/Students/sharma/theses/Shr05MS.pdf>
- [11] D. Bhatia, "Approaches for validating frequent episodes based on periodicity in time-series data," Master's thesis, The University of Texas at Arlington, 2005. [Online]. Available: <http://itlab.uta.edu/ITLABWEB/Students/sharma/theses/Bha05MS.pdf>
- [12] S. Savla, "Efficient Main Memory Algorithms for Significant Intervals and Frequent Episode Discovery." Master's thesis, The University of Texas at Arlington, December 2006.
- [13] S. Savla and S. Chakravarthy, "Efficient Main Memory Algorithms for Significant Episode Discovery," *Accepted to the International Journal of Data warehousing and Mining*, vol. 3, no. 3, 2007.
- [14] A. Srinivasan, D. Bhatia, and S. Chakravarthy, "Discovery of Interesting Episodes in Sequence Data." in *Proceedings, Annual ACM Symposium on Applied Computing (SAC)*, Dijon, France, 2006, pp. 598–602.
- [15] S. Savla and S. Chakravarthy, "A Single Pass Algorithm for Detecting Significant intervals in Time-Series Data." in *Proceedings, ADBIS Workshop on Data Mining and Knowledge Discovery (ADMKD)*, Thessaloniki, Greece, 2006, pp. 49–60.
- [16] S. Shrestha and S. Chakravarthy, "SQL-Based Approach to Significant interval Discovery in Time-Series Data." in *Proceedings, ADBIS Workshop on Data Mining and Knowledge Discovery (ADMKD)*, Thessaloniki, Greece, 2006, pp. 73–86.
- [17] A. Srinivasan, S. Shrestha, and S. Chakravarthy, "Discovery of significant intervals in sequential data." in *Proceedings, ADBIS Workshop on Data Mining and Knowledge Discovery (ADMKD)*, Sept. 2005, pp. 87–98.
- [18] A. Srinivasan, D. Bhatia, and S. Chakravarthy, "Discovery of interesting episodes in sequence data," in *Proc. of the 21st ACM Symposium on Applied Computing*, France, April 2006, pp. 599–602.
- [19] S. Padmanabhan, "HDB-Subdue: A Relational Database Approach to Graph Mining and Hierarchical Reduction." Master's thesis, The University of Texas at Arlington, December 2005.
- [20] R. Balachandran, S. Padmanabhan, and S. Chakravarthy, "Enhanced DB-Subdue: Supporting Subtle Aspects of Graph Mining Using a Relational Approach." in *Proceedings, Pacific-Asia Conference on Data Mining and Knowledge Discovery (PAKDD)*, Singapore, 2006, pp. 673–678.

- [21] S. Pradhan, “A Relational Database Approach to Frequent Subgraph (FSG) Mining,” Master’s thesis, The University of Texas at Arlington, August 2006.
- [22] R. Adaikkalavan, “Generalization and Enforcement of Role-Based Access Control using a Novel Event-based Approach.” Ph.D. dissertation, The University of Texas at Arlington, August 2006.
- [23] R. Adaikkalavan and S. Chakravarthy, “How to Use Events and Rules for Supporting Role-Based Security? (Invited Paper).” in *DEXA Workshops (Invited Paper)*, Krakow, Poland, 2006, pp. 698–702.
- [24] —, “Discovery Based Role Activation in RBAC.” in *Workshop on Information Assurance, in conjunction with IEEE International Performance Computing and Communication Conference*, Phoenix, Arizona, 2006.
- [25] C. H. H. Subramanian, “Adaptive Load Balancing and Change Visualization for WebVigiL.” Master’s thesis, The University of Texas at Arlington, December 2006.
- [26] A. Sanka, S. Chamakura, and S. Chakravarthy, “A Dataflow Approach To Efficient Change Detection of HTML/XML Documents in WebVigiL.” *Computer Networks*, vol. 50, no. 10, pp. 1547–1563, 2006.
- [27] V. Garg, “Estream: An Integration and Event and Stream Processing,” Master’s thesis, The University of Texas at Arlington, December 2005.
- [28] B. Kendai, “Runtime Optimization and Load Shedding in MavStream: Design and Implementation.” Master’s thesis, The University of Texas at Arlington, December 2006.
- [29] V. Garg, R. Adaikkalavan, and S. Chakravarthy, “Extensions to Stream Processing Architecture for Supporting Event Processing.” in *Proceedings, International Conference on Database and Expert System Applications (DEXA)*, Krakow, Poland, 2006, pp. 945–955.
- [30] A. Gilani, S. Sonune, B. Kendai, and S. Chakravarthy, “The Anatomy of a Stream Processing System.” in *Proceedings, British National Conference on Databases (BNCOD)*, Belfast, Northern Ireland, UK, 2006, pp. 232–239.
- [31] S. Chakravarthy and V. Pajjuri, “Scheduling Strategies and Their Evaluation in a Data Stream Management System.” in *Proceedings, British National Conference on Databases (BNCOD)*, Belfast, Northern Ireland, UK, 2006, pp. 220–231.
- [32] S. R. Varakala and S. Chakravarthy, “Dynamic Programming Environment for Active Rules.” in *Baltic Conference on Database and Information Systems*, Vilnius, Lithuania, 2006, pp. 3–16.
- [33] Q. Jiang and S. Chakravarthy, “Anatomy of a Data Stream Management System.” in *Proceedings, International Conference on Advanced Database and Information Systems (ADBIS)*, Thessaloniki, Greece, 2006, pp. 128–143.
- [34] S. Chakravarthy, R. Dasari, S. R. Varkala, and R. Adaikkalavan, *Events and Rules for Java: Using a Seamless and Dynamic Approach*. IOS Press, 2006, vol. 155, pp. 3–17.
- [35] A. Telang, R. Mishra, and S. Chakravarthy, “Ranking Issues in Information Integration.” in *ICDE Workshop (DBRank)*, Istanbul, Turkey, 2007, pp. 257–260.

## 6 Mobile Ad-Hoc Networks

*Faculty Involved:* Mukhesh Singhal, Raphael Finkel, Sajal Das

*Student Members:* Venkata Giruka (PhD completed 2007), Huaizhi Li (PhD completed 2006), Rendong Bai, and Saikat Chakrabarti

### Major Research and Education Activities of the Project

The main objective of the project is to design and evaluate protocols for secure and efficient routing among a group of users in pervasive and mobile computing systems. The main objective is to insure secure information exchange among a group of users communicating over a wireless network. Research activities focused on investigating several key problems in insuring secure and efficient information exchange among a group of users communicating in ad hoc networks and mobile computing environments. This project provided training opportunities to four Ph.D students. The students made significant progress towards

their Ph.D degrees while supported as RA on the project. The students learned to conduct research and evaluate the developed techniques. They were trained in the important area of the design of secure wireless networks.

In addition, the research results were incorporated into the curriculum of a graduate level in secure networking course taught at the University. Students taking these courses were exposed to cutting edge technologies related to the area of the project.

## Major Findings Resulting from these Activities

- **Nodes cooperation:** Nodes cooperation is essential for proper functioning of ad hoc networks. To encourage nodes cooperation, selfish nodes are reimbursed for their individual forwarding cost, necessitating the need for truthful protocols to prevent these nodes from cheating on their true cost. Meanwhile, it's of same importance for such truthful protocols to achieve good network performance.

We developed a Low Overhead Truthful routing protocol for route discovery in MANETs with selfish nodes by applying mechanism design. This protocol produces an overhead of  $O(n^2)$ , a significant improvement over the existing solutions'  $O(n^3)$ , and achieves much better network performance than those protocols. We further developed a Light-weight Scalable Truthful routing Protocol (LSTOP), which approaches the optimal routing in dense networks. The most prominent feature of LSTOP is that it causes a significantly low message overhead of  $O(n)$  on average, thus it is much more scalable.

- **Security in Position-based Protocols:** Apart from the usual attacks on routing protocols such as modification, dropping, black hole, etc., position-based protocols face a new attack, viz., the position spoofing attack where a node declares a fake position and utilizes it to launch other attacks.

We developed a secure position-based protocol framework, S-CORE, for ad-hoc networks. S-CORE consist of a secure hello protocol and a secure greedy forwarding protocol (SecFwd). The secure hello protocol helps nodes establish a neighbor table free of malicious nodes, defends against ID-spoofing, position-spoofing, and black lists malicious nodes. To detect position-spoofing, nodes use a new distributed position verification algorithm that uses a position-bounding technique. SecFwd mitigates forwarding misbehaviors and insures proper execution of greedy forwarding. We presented a security analysis of S-CORE and results from an extensive simulation study using GloMoSim. Our simulation results showed that S-CORE detects misbehaving nodes to a high degree with a fewer number of false accusations.

- **Adaptive Protocols for Pervasive and Mobile Systems:** We developed a localized IP auto-configuration protocol for wireless ad-hoc networks, called *OASIS*. *OASIS* introduces the idea of position-dependent IP-addresses using a cellular structure of the underlying network. Within a cell, nodes try to obtain a conflict-free IP-address using a distributed mutual exclusion style algorithm. We showed the correctness, and present an asymptotic analysis of the overhead of *OASIS*. Further, we presented extensions to *OASIS* protocol to cope with message losses and node mobility. *OASIS* avoids problems due to network partitions and mergers, supports concurrent node joins/leaves, incurs localized control overhead per IP-address acquisition (independent of the number of nodes in the network), and hence it is scalable.

We also developed a position-based protocol framework that copes with both selfish nodes and unidirectional links. To cope with selfish nodes, we present a mechanism to stimulate cooperation among nodes. The basic idea is that nodes that forward packets for a source-destination pair get payment (and even some bonus) for their forwarding service from the source or the destination. However, the challenge in using this mechanism is that nodes utilizing forwarding service may want to pay a lowest possible payment to nodes offering the forwarding services, and selfish nodes may not reveal their true cost to maximize their payoff. This necessitates the need for designing a *truthful* protocol. A protocol is *truthful* (or strategy-proof) if it maximizes the nodes' payoff only when they reveal their true cost.

To solve the problem, we developed an Auction-Based Greedy Forwarding Scheme (AgFS) for ad-hoc networks. AgFS introduces an auction-based packet forwarding scheme that *guarantees truthfulness* while inheriting the localized nature of greedy forwarding. We proved that AgFS is truthful, and we statistically analyzed the average progress made per hop using AgFS.

To cope with unidirectional links, we developed an enhanced version of the hello protocol which detects and avoids using unidirectional links. The reason for avoiding unidirectional links is explained in detailed in the next section. Further, we combined the hello protocol and AgFS into a single framework. We used this framework in conjunction with the OGPR to develop a robust position-based protocol, namely, OGPRv2, for ad-hoc networks. To the best of our knowledge, OGPRv2 is the first protocol of its kind in the context of position-based protocols for ad-hoc networks.

We deployed a six-node ad-hoc network test-bed at the University of Kentucky and developed an OGPRv2 prototype. The project involved an MS thesis student and a Ph.D. student as a research assistant.

- **Way Point Routing Model (WPR) and DSR Over AODV Routing Protocol (DOA):** We developed a lightweight hierarchical routing model called Way Point Routing (WPR) for MANETs. WPR selects a number of nodes on a route as waypoints and divides the route into segments at the waypoints. Waypoints run a high-level inter-segment routing protocol; nodes on each segment run a low-level intra-segment routing protocol. When an intermediate node moves out or fails, previous protocols discard the whole original route and discover a new route, but WPR only requires the two waypoints of the broken segment to repair that segment.

In addition, we developed an instantiation of WPR, in which we use DSR as the inter-segment routing protocol and AODV as the intra-segment routing protocol. We call this instantiation the DSR Over AODV (DOA) routing protocol. DOA combines DSR and AODV, two well-known on-demand routing protocols, in a hierarchical manner and includes each of these protocols as special cases when the segment size is minimal or maximal. Furthermore, we designed two novel techniques that apply to DOA: an efficient loop-detection method and multi-target route discovery.
- **Carpooling in MANETs: The Case of Multiple-Target Route Discovery:** Carpooling is an efficient transportation option in real life because it reduces traffic and cost. We applied this idea to MANETs and developed Multiple-Target Route Discovery (MTRD). MTRD aggregates multiple route requests into one RREQ message and discovers multiple targets simultaneously. When a node has to find a route to a destination, instead of immediately injecting a new RREQ message into the network, the node tries to discover a route by attaching its request to in-transit RREQ packets that it relays for other nodes. MTRD improves routing performance by reducing the number of regular route discoveries. We also analyzed several important properties of MTRD.
- **Route Discovery in MANETs: From Unilaterality to Bilaterality:** Traditionally, route discovery in MANETs operates in a source-initiated manner. This unilateral operation is unbalanced and inherently not ideal for MANETs. We designed a new scheme called Bilateral Route Discovery (BRD), in which both source and destination actively participate in the process of discovering a route between them. BRD has the potential to halve the control overhead. As an underlying protocol for BRD, we developed Gratuitous Route Error Reporting (GRER). GRER bypasses a failed link and notifies the destination of a broken route. The destination can thus play an active role in the ensuing route re-discovery.
- **Salvaging Route Reply for On-Demand Routing Protocols:** The research community prefers on-demand routing protocols in networks like MANETs, in which resources such as energy and bandwidth are constrained. In these protocols, a source typically discovers a route to a destination by flooding the network with a route-request (RREQ) message and waiting for a route-reply (RREP) message. A RREP travels hop by hop towards the source. Sometimes an intermediate node can not send the RREP to the intended next hop due to changing topology or congestion. Existing routing protocols discard the undeliverable RREP. This loss is highly unacceptable because a RREP message has a lot at stake: It is obtained by tens or hundreds of RREQ transmissions, which is an expensive and time-consuming process.

We developed the Salvaging Route Reply (SRR) algorithm to address this problem. SRR has two schemes: SRR1 and SRR2. In SRR1, a node salvages a RREP by broadcasting a one-hop salvage request. In SRR2, a node uses duplicate RREQs to maintain a backup path to the source. SRR2 is an improvement over SRR1 because SRR2 does not incur extra control messages.

## References

- [1] Rendong Bai and Mukesh Singhal, "DOA: DSR Over AODV Routing for Mobile Ad-Hoc Networks," *IEEE Transactions on Mobile Computing*, Volume 5, Issue 10 (October 2006) Pages: 1403-1416.
- [2] Rendong Bai and Mukesh Singhal, "Carpooling in Mobile Ad Hoc Networks: the Case of Multiple-Target Route Discovery," *WiOpt 2007: 5th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, accepted, Limassol, Cyprus, April, 2007.
- [3] Rendong Bai and Mukesh Singhal, "BRD: Bilateral Route Discovery in Mobile Ad Hoc Networks," to appear in *IFIP Networking 2007: the sixth International Conferences on Networking*, Atlanta, Georgia, May 2007.
- [4] Rendong Bai and Mukesh Singhal, "Enhancing Performance by Salvaging Route Reply Messages in On-Demand Routing Protocols for MANETs," submitted to *Ad Hoc & Sensor Wireless Networks*, 2007.
- [5] Rendong Bai and Mukesh Singhal, "Route Discovery in Mobile Ad Hoc Networks: From Unilaterality to Bilaterality," to appear in *Mobiquitous 2007: the 4th Annual International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, Philadelphia, PA.
- [6] Venkata C. Giruka Mukesh Singhal, "A Localized IP-address Auto-configuration Protocol for Wireless Ad-hoc Networks", *ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots (WMASH) 2006*. In conjunction with *ACM MobiCom September 2006 Los Angeles, California, USA*, pp. 101-108.
- [7] Venkata C. Giruka, Yongwei Wang, and Mukesh Singhal, "A Secure Position-based Protocol Framework for Wireless Multi-hop Networks", submitted to *IEEE WiMob 2007*.
- [8] Venkata C. Giruka, Yongwei Wang, Saikiran Madgula, and Mukesh Singhal "Position-based Routing Under Non-Ideal Environments", in preparation.
- [9] Venkata C. Giruka and Mukesh Singhal "A Self-Healing On-Demand Geographic-Path Routing Protocol for Mobile Ad-hoc Networks", accepted to be published in *Ad Hoc Networks*, Elsevier.
- [10] Yongwei Wang and Mukesh Singhal, "On Improving the Efficiency of Truthful Routing in MANETs with Selfish Nodes", accepted by *Journal of Pervasive and Mobile Computing*, Elsevier.
- [11] Yongwei Wang and Mukesh Singhal, "LSTOP: A Light-weight Scalable Truthful Routing Protocol in MANETs with Selfish Nodes", In *Proceedings of 5th International Conference of Ad-Hoc, Mobile, and Wireless Networks (ADHOC-NOW) 2006, Ottawa, Canada, Aug 2006, (Springer LNCS 4104)*, pages 280-293.
- [12] Yongwei Wang and Mukesh Singhal, "A Light-weight Scalable Truthful Routing Protocol in MANETs with Selfish Nodes", submitted to *International Journal of Ad Hoc and Ubiquitous Computing*.
- [13] Yongwei Wang, Venkata C. Giruka and Mukesh Singhal, "Truthful Multipath Routing for Ad-hoc Networks with Selfish Nodes", submitted to *Journal of Parallel and Distributed Computing*.
- [14] Venkata C. Giruka, Yongwei Wang and Mukesh Singhal, "Position-based Routing Under Non-ideal Environments", Work in progress.
- [15] Venkata C. Giruka, Yongwei Wang and Mukesh Singhal, "A Secure Position-based Protocol Framework for Wireless Multi-hop Networks", submitted to *WiMob'07*.
- [16] Yan Sun, Qiangfeng Jiang, Mukesh Singhal, "An Edge Constrained Localized Delaunay Graph for Geographic Routing in Mobile Ad Hoc Networks", In *proceedings of IEEE WCNC'07*.

## 7 Direct Impact of PSI

### Project Coordination and Interaction

The PSI project has fostered excellent working relation and collaboration among the PIs and Co-PIs from UTA, UKY and PSU. There is intense exchange of ideas between the various personnel so as to

develop an understanding of the mechanism and architecture on how to transform useful research into meaningful working prototype.

At UTA, there are regular monthly face-to-face meetings with students and faculty working in various project groups and subgroups. Two workshops (2004, 2005) on PSI have already been successfully conducted at UTA. The workshops provide a platform that encourage the faculty and students to share research findings, ideas, and discuss plans for further collaborative research.

### **Project Outcome - To Date**

The PSI project has already resulted in ten PhD dissertations from the University of Texas at Arlington (UTA), two from Pennsylvania State University (PSU), and two from the University of Kentucky. At the same time, fourteen Masters theses were awarded from UTA, two from PSU, and ten from the University of Kentucky. To date, there are about twenty PhD dissertations being pursued at UTA. This project has supported five PhD students at UKY and five PhD students at PSU. Two undergrad students have been trained through REU supplements at UTA. Four new graduate level courses on have been introduced at the three collaborating universities. In particular, one graduate course on Wireless Security and another on Advances in Sensor Networking have been designed and taught at UTA in Spring 2005; a graduate course on Wireless Networks Security was taught at UKY in Fall 2004; and an advanced graduate level course in the area of Heterogeneous and Mobile Data Bases was developed and offered at PSU in Spring 2005. These courses will be offered at regular intervals in respective universities. Furthermore, the course materials will be shared and offered at the collaborative universities. The objectives of such courses include the dissemination of basic and advanced concepts in those emerging topics, and encouraging research among students. Students are gaining hand-on experience in experimenting with sensor network testbed developed at UTA. Finally, students have had the opportunity to publish their research work at high quality conferences and journals.

### **Project Leverage**

Multiple security projects spun out of this ITR project. They include SafetyNet: project for border security including perimeter control, airport/harbor security, Content Based Routing for imemdiated notification of security events, High performance packet classifier for anomaly detection in traffic streams, and video sensor networks for surveillance in highly uncertain dynamic environments. They were submitted to NSF, Federal Earmark funding and Intel Corporation. Additionally, there are two proposals pending with the Cybertrust program and with NOSS.

A new research proposal aimed, at building a multi-layer security framework, got funded in May 2006 by the Texas Advanced Research Program. Additionally, an NSF MRI grant for training undergraduates in network and pervasive security also got funded in 2006. This is in addition to the NSF MRI proposal that got funded in 2004 to help procure equipment needed to develop a prototype of the reseach proposed in this ITR project. This ITR project has also resulted in a spin-off in Bioinformatics and System Biology. This has further led to collaboration with UT Southwestern Medical center in Dallas.