Summary

In this proposal entitled **Modular Strategies for Internetwork Monitoring** we address the long-standing and difficult problem of detecting and classifying spatially distributed network anomalies from multiple monitoring sites. Characterizing baseline vs. anomalous behavior of the Internet requires deployment of collaborative data collection, anomaly detection and pattern recognition on a large scale. Progress on such a tough problem demands a broadbased and innovative approach which accounts for the practical constraints on privacy and information sharing, communications, and distributed data processing among the Internet's heterogeneous, and perhaps competing, network administrators and ISP's. We combine the forces of leading researchers in three complementary disciplines: (i) networking and data collection; (ii) statistical data analysis and signal processing; (iii) decentralized decision-making and discrete event systems, to develop new methods for monitoring the network which are both modular and scalable. Our effort will lead to a better understanding of the fundamental theoretical limitations impeding accurate detection in a networked monitoring environment. It will also lead to new algorithms for detection of coordinated intrusions, distributed denial of service attacks, and quality-of-service degradations.

The high-dimension and complexity of nominal packet-level and traffic-level patterns in the Internet makes establishing a baseline for anomaly detection extremely challenging. Furthermore, in administered networks, there is a natural aversion to information sharing which could impinge on privacy or erode competitive advantage. Thus any viable approach must compromise the individual good for the societal good to have any chance of being widely accepted. We will explore practical data sharing protocols which operate in conjunction with decentralized data analysis algorithms. We will also study the fundamental tradeoff between ensuring data privacy and anomaly detectability using game theory.

For our research we will adopt a modular global monitoring structure that is decomposed into a three level hierarchy: local level measurement of data from servers, routers and switches; intermediate level data analysis and processing of end-to-end traffic measurements, summary statistics and alarms transmitted from the local level; and upper level decision-making and processing of information transmitted from the intermediate level. This modular structure allows our approach to be scalable to large networks of monitoring sites. However, this structure also imposes interesting constraints on the data analysis which requires development of new approaches. Three approaches will be pursued: distributed spatio-temporal data analysis using wavelets over graphs; event detection and classification using distributed pattern analysis and learning; and multi-site event correlation using discrete event dynamical systems. The adaptation of these approaches to the scalable modular processing structure will constitute a major advance in the theory of distributed statistical information processing. Our approach goes well beyond previously introduced techniques of fault detection, traffic analysis, and alert correlation that have been restricted to much smaller scale problems.

The **intellectual merit and impact** of this proposal include: (i) the development of a general theory of distributed data analysis and anomaly detection for large-scale networks of monitoring devices subject to privacy constraints; (ii) creation of a repository of real multi-site traces of Internet anomalies which will be accessible to other networking researchers; (iii) cross fertilization to other applications of distributed spatio-temporal analysis, e.g., wireless sensor nets; networked biosensors; survey sampling for population dynamics.

The **broader impact** of this proposal include: (i) involvement of female or under-represented minority graduate and undergraduate students; (ii) outreach to Middle Schools and High Schools (grades 6-12) by a combination of presentations at local schools, participation in a summer camp emphasizing computing security, and involvement in workshops for K-12 teaching and computing staff; (iii) interaction with networking industry to help transition our research to the operator and user communities; (iv) active participation of the students involved in this project in national and international professional meetings; (v) active collaborations and student exchanges with international researchers and institutions in Canada and France.

1. Introduction and Executive Summary

Modern society increasingly relies on private and public communication networks of data terminals, sensors, routers, and switches. Consequently this crucial infrastructure must be robust to both key component failure and malevolent behavior. The original design principles of data networks did address robustness issues, but the focus was on localized link and switch failure. Now that the *internetwork* of thousands of autonomous networks interact with and depend on each other, vulnerabilities have been exposed and it has become evident that there are identifiable, isolated systems in the Internet whose failure could cripple multiple communication systems. Networks are also exposed to sophisticated attacks that target weaknesses in communication protocols, switch operation, and host software. There is thus a pressing need to fortify networks, to develop mechanisms for detecting, localizing, and classifying failure modes, and to take action to address malfunction. An extensive network monitoring infrastructure that tracks performance and can detect anomalous network behavior is required to address these needs.

We are a multi-disciplinary team of researchers from the broad areas of network measurement, statistical signal processing, and dynamical systems. With our combined strengths in data collection, data analysis, and decentralized decision-making we propose to develop a modular framework for internetwork monitoring which will result in scalable techniques for anomaly detection and a better understanding of the fundamental limits on detection performance under the practical constraints of privacy. Our approach has the following features:

1. A comprehensive plan for multi-site data collection and information sharing which leverages on existing infrastructure developed by us and collaborators at University of Wisconsin (UW), Boston University (BU), University of Michigan (UM), Arbor Networks, Merit Network, and Internet2.

2. Development of a new class of data analysis methods which operate within the communication and processing constraints of a modular decentralized information sharing structure, and which capture the tradeoff between proprietary concerns (privacy) and accurate global anomaly detection.

3. Introduction of novel network measurement and statistical inference methods that can estimate network performance in regions where monitors cannot be deployed.

4. A discrete event dynamical system (DES) framework for capturing anomalies using compact finite-state automata and semi-Markov chain state space models.

The proposal is organized as follows. After describing prior NSF support below, Section 2 discusses the background for our research. Section 3 describes our overall research approach. Sections 4, 5 and 6 describe our proposed research in more detail, which is divided into sections on Data Collection and Information Sharing (Sec. 4), Distributed Data Analysis (Sec. 5), and Discrete Event System Models (Sec. 6). Section 7 outlines our plans in education and Section 8 describes the impact of the project.

Prior NSF Support

1. Information theoretic analysis of tomographic systems, NSF BCS-9024370 (1993-1995), A.O. Hero (PI), University of Michigan: We established new criteria for design of tomographic data collection systems and developed new high performance algorithms for 2D and 3D reconstruction from projections [59, 62, 58, 44, 43, 57]. This work was a stepping stone to recent work on inference of network behavior from multi-site network data [140, 138, 139].

2. Failure Diagnosis of Modular and Decentralized Discrete Event Systems, NSF ECS-0080406 (2000 to 2003), S. Lafortune (PI) and D. Teneketzis (Co-PI), University of Michigan: The overall objective of this project is to develop a comprehensive methodology for failure diagnosis of large-scale complex systems using DES [82]. Our current research and results to-date include: (i) diagnosis of intermittent failures in the context of centralized architectures [23]; (ii) dealing with communication delays in the context of coordinated decentralized architectures [38]; (iii) failure diagnosis of stochastic automata [148]; (iv) sensor selection for failure diagnosis [37, 169]; and (v) study of the computational complexity of diagnosability

[170].

3. Commonwealth Scalable Web Servers, NSF EIA-9706685 (1997-2000) M. Crovella (PI) and A. Bestavros (co-PI), and NSF CAREER CCR-9501822, (1995-1998) M. Crovella (PI). These grants made advances the following areas: atatistical analysis of network data [29, 30, 105, 31, 106, 107, 45]; tools for on-line link condition measurement [16]; gauging the infrastructure size needed for network measurements [6, 83]; spatial characterization of network traffic [28].

4 A Multiscale Framework for Spatial Modeling in Geography, BCS-0079077, 2000-2003, E. Kolaczyk (PI) and Sucharita Gopal (Co-PI), Boston University: The focus of this work has been on the development of non-traditional multiscale statistical modeling frameworks for spatial data [76], the exploration of their nature and properties [86, 74, 75], and their application to specific problem areas in geography [51] and remote sensing [68, 69, 73]. Contributions to human resources include the mentoring of three graduate students and two undergraduate students in mathematics/statistics and geography, with the latter supported by an REU supplement to the original grant.

5. INCITE: A Framework and Methodology for Edge-Based Traffic Processing and Service Inference, NSF ANI-0099148, (2001-2004), R. Nowak (PI), E. Knightly (co-PI), R. Baraniuk (co-PI), R. Riedi (co-PI), Rice University: The goal is to indirectly infer dynamic network characteristics using only edge-based network traffic processing, without special-purpose network support. A large number of conference and journal articles have resulted from this project to date, which can be found at the website `http://spin.rice.edu/NSF`. The project has involved several undergraduate students (via the NSF-REU program); several of whom have co-authored papers appearing at top IEEE and ACM conferences.

6. Instrumented Streaming Research and Testbed, NSF ANI-0117810 (2001-2004) M. Vernon (PI) and P. Barford (co-PI) University of Wisconsin-Madison. This project encompasses the design, implementation, and state-of-the-art instrumentation of new methods for scalable wide-area on-demand reliable digital (SWORD) streaming. The larger goal of the research is to enable new streaming media applications, such as immediate access to an arbitrary television show or other stored media content whenever a client anywhere would like to view it [5, 127, 50, 2, 87]

2. Background

The need for a hierarchical, distributed approach to data collection and anomaly detection can be seen by considering the nature of typical anomaly detection problems. Clearly, when traffic exhibits unusual characteristics, an immediate and fundamental question concerns the size and extent of the region over which the anomaly occurs. For example, if observed traffic load increases to an unusual level, this may be due to a number of factors. Traffic throughout the network may have risen due to some external driver of increased demand such as a breaking news story. Alternatively, traffic in a localized network region may be increased due to a flash crowd effect (publication of a popular video or report that drives traffic to a single location). Finally, traffic load may be due to a particular pair of hosts engaging in abnormally high traffic. These three scenarios are primarily distinguished by the size of the "neighborhood" over which the anomalously high traffic is observed, and they each demand a different response from network operators.

Anomalous network conditions may arise due to malicious behavior (attacks), or due to effects stemming from network operations. In each case, a hierarchical, distributed approach to problem assessment is needed. As an example of this need in assessing network attacks, consider the problem of rapidly detecting denial of service (DoS) attacks. This capability is crucial for responsive network management. Unfortunately, increased traffic on a single link is not a good indicator of the presence or nature of a DoS attack. Most DoS attacks are distributed, with flooding packets arriving from multiple sources along multiple paths. Accurate identification of a distributed DoS attack using traffic counts requires the simultaneous assessment of traffic on multiple links of the network.

Turning to the problem of assessing anomalies in network operations, it is clear that most issues that arise are fundamentally distributed and multi-scale, as illustrated in the next four paragraphs.

First, consider the need to understand shifts in traffic patterns as a result of network equipment performance degradation or failure. For example, some networks are engineered with sufficient bandwidth for "protection," i.e., so that traffic shifts due to equipment failures can be absorbed without manual intervention in the routing system. In other cases, networks are provisioned with the expectation that equipment failures will be addressed through explicit traffic engineering actions. In each case, it is essential to have a whole-network view of how traffic patterns shift when equipment fails or traffic is manually re-routed. This whole-network view must provide quantitative information about which regions of the network experienced increased load and which experienced decreased load as a result of the network event.

A second such problem is the detection of routing loops. Routing loops are notoriously hard to detect in networks due to the lack of adequate, efficient tools. As a result anecdotal evidence suggests that some routing loops persist for very long periods of time. When a routing loop develops, individual packets may traverse the loop hundreds of times before being dropped from the network. This can increase the load on the particular set of links by a considerable amount. In fact, routing loops should be detectable due to this increased traffic, but such detection requires the comparison of traffic counts over a large set of varying regions of the network.

A third example of an operational problem can be seen to arise in the Border Gateway Protocol (BGP) system. Recent work [24] has identified the presence of BGP "storms", periods when BGP traffic can be 3 to 4 orders of magnitude higher than normal, lasting for hours. Precisely assessing the extent of such storms based on traffic flows alone is beyond the capability of current tools and methods. No existing tools can summarize traffic measurements over a range of topological neighborhoods. Studies to date have only managed to explain a small set of examples of traffic storms; progress toward a more systematic, statistically based study is critically dependent on the development of more powerful tools.

Finally, there are a number of problems that are associated with individual links as well. Many such problems fall under the general heading of traffic variability. Variability or instability in traffic flows is a considerable concern for network operators. In many cases, such variability is due to characteristics of applications sharing the links, but in other cases it may be due to hardware instability. For example, periodicity or increased burstiness in traffic on a given link may reflect route flapping, or periodic interruptions of service in some router.

When traffic properties change in an important way—e.g., when traffic becomes more variable or bursty at fine time scales—it may signal the onset of system instability. This is an important event that may need attention. Unfortunately, tools for identifying such changes in short-term variability are lacking. The vast majority of tools available for traffic analysis concentrate on time-invariant or long-timescale properties (e.g., long range dependence). While such properties are important, they do not inform network engineers about any special aspects of the current conditions on the network. Furthermore, to identify likely sources of traffic variability or changes, it would be helpful to be able to correlate traffic changes across links. For example, if some set of links in sequence show similar periodic behavior, then the application responsible may be associated with the endpoints of the multi-link path.

To make progress on these challenging problems for large scale networks will require advances in both data collection infrastructure and data analysis.

3. Research Approach

It is clear from the discussion in Section 2 that to deal with anomaly detection in large-scale networks we need to adopt a hierarchical and distributed approach. The modular information processing and data collection architecture that we adopt strikes a natural compromise between a fully centralized and a fully decentralized architecture. While a fully centralized architecture would certainly offer the best anomaly detection performance, it is infeasible due to the following reasons: (a) the amount of raw data generated at the monitoring sites (traffic, packet, and host level) is enormous; (b) the links between monitoring and processing sites are of limited bandwidth for transmitting this data; (c) statistical analysis and modelling of

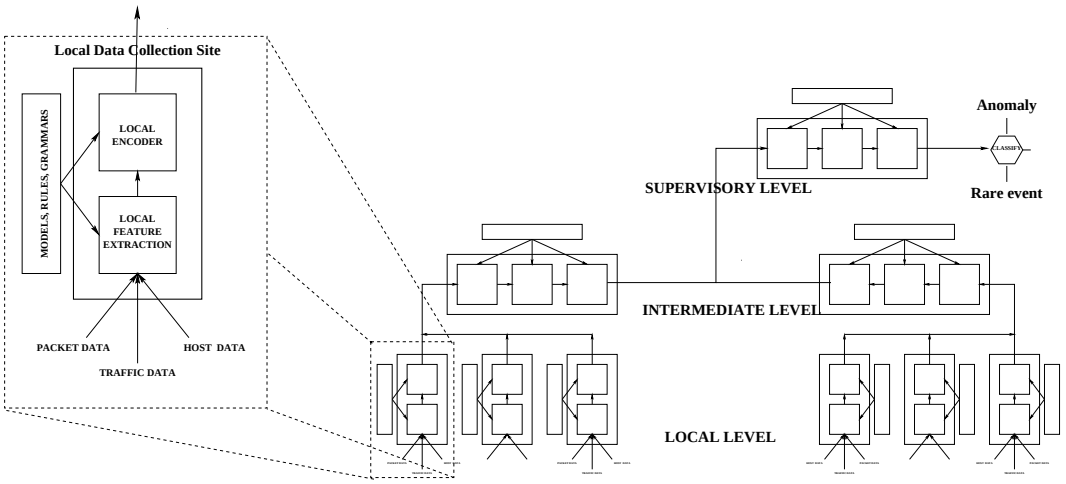

Figure 1: *Block diagram of modular architecture for data collection, information processing, and decision making.*

such a huge amount of data is impractical; and, perhaps most importantly, (d) information privacy becomes the overriding concern when a single site has access to all raw data. On the other hand, a fully decentralized architecture offers the best information privacy but is inadequate due to the following reasons: (a) single sites are incapable of detecting multi-site correlations which are necessary for identifying distributed attacks on the network; and (b) information sharing between neighboring sites is necessary to establish a baseline of normal network operation.

For the above reasons we believe that a modular approach (that will be explained below) is the most promising method for anomaly detection in large-scale networks. This architecture is motivated by (i) the natural way that data collection is done in the network, and (ii) the need to accommodate data privacy considerations. The modular architecture we propose is illustrated in Figure 1 and consists of a three level hierarchy. Each level has the following functionality. The local level consists of individual or small groups of routers, sniffers, or servers, that acquire raw measurements such as host data (user commands, ftp/http processes), packet headers (SYN, OD, ACK), and traffic streams (Netflow, end-end delay/loss tomography). Extracted features are encoded at the local level and sent to a domain server (e.g., an Automous System (AS) or its agent) at the intermediate level for aggregation and processing. Finally, intermediate decisions and other aggregated information are communicated to a supervisory level which serves several domains. Within this architecture each level can make decisions concerning attacks and anomalies in its own domain.

Within the context of this architecture our principal aim is to develop implementable modular monitoring strategies which can deal with a mixture of continuous-state data, e.g., measured traffic flows and end-to-end delays, and discrete-state data, e.g., `tcpdmp` fields and other timed event sequences. We will develop detection algorithms and investigate their performance for the proposed architecture using a combination of model-based and data-driven approaches (described in more detail in Sections 5 and 6). Data-driven analysis using a hierarchy of spatio-temporal wavelets and kernel-based classifiers will be used to explore baseline vs. anomalous behavior of network flows and other data. Results of this analysis will be used to develop and validate a class of novel distributed spatio-temporal discrete event dynamical system (DES) models for real-time tracking of changes from baseline. The DES models at any level of the hierarchy will be driven by discrete event signals that are either transmitted from other levels or are derived from a combination of discrete- and continuous valued data measured at that level.

4. Distributed Data Collection and Information Sharing

Our near-term objective is to gather and archive Internet anomaly data for the purposes of developing detection methods and for limited distribution within the research community. Our long term objective for

distributed anomaly data collection is to create a framework for this activity that can be used throughout the Internet. There would be many advantages in such a system. For example, consider the problem of identifying sources of denial-of-service (DoS) attacks. Most of the currently proposed methods for trace-back require some kind of support from the routing infrastructure. If a DoS attack were launched while distributed anomaly monitoring were in use, then participating nodes could identify offending flows in their networks and create a trace tree from target to sources. In this section we describe our proposed approaches for data collection and information sharing.

4.1. Data Collection and Archival

An important component of this proposal is the collection of anomaly and attack data from sites around the Internet. This data will form the foundation for the development of the detection strategies described in later sections. Data collection will be facilitated through the deployment of new measurement and monitoring tools in a variety of existing widely deployed infrastructures. Data generated by these systems will be gathered and stored in a new Internet attack and anomaly archive. This system will fuse measurements from the multiple sites to provide a detailed, consistent dataset for detection strategy development. Our efforts will leverage the combined forensic expertise of our collaborators at Merit Network, Internet2, Arbor Networks and co-PI Barford to identify and annotate interesting events (downed links, storms, attacks, etc) which will be targetted for inclusion in our anomaly archive.

4.1.1. Goals and Challenges in Data Collection and Archival: Our challenge is in addressing the general problem of balancing the quantity of data collected with what is required to precisely identify anomalies. Our hypothesis is that coordinating measurements from multiple sites improves precision of anomaly identification. However, this must be done in a way that does not have a serious negative impact the network or network systems.

The traffic monitoring systems will provide data to a centralized traffic repository which we call the Internet Attack and Anomaly Archive (IA3). The IA3 will maintain the measurement data in a format that is accessible by both the team and (eventually) the network community at large. It will also provide a front end for extraction and evaluation that enables a consistent, secure perspective in an environment that balances privacy with access to data required to identify anomalies.

To develop an understanding of these data collection and archival issues we plan to investigate the following questions:

1. How does sampling in packet and flow level measurements affect the ability to detect anomalies?

2. Can data collected at multiple sites be merged in a way that reduces the overall volume of data without affecting anomaly detection ability?

3. What is the extent to which measurement resolution can be improved through coordination between multiple sites?

4. What is the architecture of a centralized traffic measurement repository that promotes participation through privacy preservation and enables effective coordinated anomaly detection?

5. What is the architecture of a distributed anomaly repository?

4.1.2. Research in Monitoring for Attacks and Anomalies: Measurement and monitoring are fundamental activities in wide area network operations. A primary focus in these activities is detecting and diagnosing significant deviations from an established baseline behavior - so called *anomalies*. Standard best practices for wide area network monitoring include the use of Simple Network Management Protocol (SNMP) [143] data and to a lesser extent, the use of IP flow data [18]. The nature of this traffic data is that it is typically quite simplistic; such as counts of packets or loss rates at a router interface. In the case of flow data, through the use of tools like FlowScan [113], application and more network specific data may be gleaned.

A fundamental step in detecting network traffic anomalies is simply establishing a definition of “anomaly.” It is not uncommon to hear a network engineer say, ”I know one when I see one”. In fact,

there is no hard-fast definition of an anomaly without additional network specific context. A first step in defining anomalies is in the creation of categories of anomaly types. Barford et. al. suggest four categories for anomalies in [4]: network (eg. failures and outages), attack (eg. standard SYN floods), flash crowd and measurement (catch-all category). Intrusions can be considered a special case of network anomalies in the sense that they are typically attempted in such a way as to avoid detection and do not result in visible changes in network traffic data. Detection of network intrusion activity (with the exception of worm outbreaks [144] which can also be considered attacks) is not a focus of this proposal.

Even with a definition of anomaly, monitoring is hard for a number of reasons. The first is the inherent variability in network traffic [84, 109, 26]. This makes any simple thresholding methods for detecting anomalies virtually useless. This also speaks directly to perhaps the single most significant problem in anomaly detection: the reduction or elimination of false positives from automated systems. The second significant challenge in monitoring is determining when, where, what and how to gather data. Many traffic characteristics can be monitored and from many points in the network. The objective is to balance quantity of data with the ability to discern anomalies. The third challenge is in overcoming the enormous logistical difficulties in deploying and maintaining widely distributed infrastructures. Issues include privacy (the overriding concern for sites considering participation), data normalization (insuring data received from multiple sites has a common representation), management (dealing with faults and failures at participating sites).

Our objective is to develop and maintain an anomaly monitoring system that will provide data for the modeling and detection strategy development efforts. Our approach will be to use and extend existing widely deployed measurement infrastructures. Infrastructures initially targeted for this project include Internet2/Abilene [1], PlanetLab [33], Surveyor [70], DSHIELD [151], and DOMINO [168]. These infrastructures all provide access to measurement and monitoring systems in either end hosts (PlanetLab, Surveyor) or within networks (Internet2, DSHIELD, DOMINO). DSHIELD and DOMINO are both monitoring infrastructures specifically focused on attacks and intrusions. A PI (Barford) runs both Surveyor and DOMINO and has direct access to DSHIELD. UW-Madison is a participant in PlanetLab thus direct access to that infrastructure is also available. Internet2/Abilene has agreed to provide access to traffic measurements. The combination of these systems provides significant distributed capability for deploying measurement and monitoring tools developed in this project.

An important component in our envisioned monitoring infrastructure is the development of an anomaly data gathering module for the FlowScan tool. FlowScan, developed at UW-Madison, is a tool for gathering, decomposing and archiving flow measurements from routers. It is currently deployed at over 300 sites world wide. We plan to develop an enhancement to FlowScan that will enable it to participate in a coordinated anomaly monitoring system. This enhancement will benefit local sites by providing data from other participants - hopefully making it easier to detect and identify anomalies. It will also provide a framework for installing detection tools developed in this project. The wide acceptance and use of FlowScan should facilitate participation by sites outside of those with whom the PIs are directly affiliated - the obvious first candidates for deployment.

Using the distributed infrastructure as a platform for data gathering, we plan to investigate methods for improving measurement precision while reducing the impact of measurement on the network. Consider as an example the problem of flow-level measurements in routers. It is well known that enabling all flow monitoring features in routers can result in a significant reduction in switching performance (as much as 40% depending on systems and their configurations). One way to deal with this is through flow sampling techniques. The problem is that sampling can reduce precision of measurement. No one to date has addressed the problem of how sampling affects the ability to detect and identify anomalies - this will be one focus of our measurement efforts.

4.1.3. Research in Anomaly and Attack Archival: Network measurement archives of any kind are few

and far between. Examples include the Internet Traffic Archive [34] (variety of data sets), the Web Traffic Repository [53] (web cache and server logs), NLANR's PAM and AIM archives [47] (packet and delay measurements), CAIDA's Skitter repository [141] (traceroute data), and Oregeon's Routeviews [125] (BGP data). At present there are no sources or repositories of network traffic anomaly data. More generally, network data repositories of any kind are difficult to develop and maintain. Paxson discusses some of the difficulties in [108]. Primary among these are the issues of consistency, perspective and privacy.

Our objective with IA3 is to develop a first-of-its-kind repository for network traffic anomalies. Issues which must be addressed in development of this system include data normalization, data compression, data privacy. We plan to address normalization problems through transformation tools developed for each of the measurement systems that will be participating in the monitoring infrastructure. Operationally, standard Lempel-Ziv lossless compression (`gzip`) will be used at each data collection site. However, one of our research aims (see research item Q1 in Section 6.2) is to investigate distributed lossy compression (source coding) schemes that are tailored to the anomaly detection architecture discussed in Section 3.

Privacy in any kind of network measurement data is a complex and subtle issue. Policies for data sharing network data vary widely but typically focus on IP addresses and payload content. We will anonymize IP addresses through prefix preserving methods similar to those discussed in [167]. Access to any aspect of content will likely be very restricted. However, access to data collected from *unused* IP addresses (such as in DOMINO) offer limited access to content. To be successful in our data sharing efforts, a PI (Barford) will have the responsibility of coordinating with each site in our measurement infrastructure to insure that local policies are enforced.

The foundation of IA3 will be in the development of a comprehensive database environment. This system will be developed through coordination with the UW-Madison database group including Professors David DeWitt and Jeffrey Naughton. A proof of concept database is currently under development for use in the Surveyor infrastructure. That system is providing invaluable insights on appropriate schema, query and management policies for dealing with the extreme size of network measurement datasets. An important component of the IA3 system will be the development of data management and retrieval systems. These systems will enable rapid loading of data from measurement systems, and a convenient access point to data for people using the repository.

The IA3 will be housed at in the Wisconsin Advanced Internet Lab. This facility currently houses the data repositories for Surveyor and DOMINO. It includes at present over 10TB of total storage capacity - 2TB of which can immediately be allocated for IA3. The bulk of the storage capacity is in two EMC Symmetrix series 3000 systems which provide extremely high speed and reliability. As the project progresses, additional storage capacity will need to added. The Symmetrix systems can each be scaled to 50TB giving the project ample room for growth. To appreciate why this growth may be needed consider that one month's worth of flow logs from all backbone routers in Abilene is approximately 2TB. The task of simply downloading this amount of data to a central site over the network can be extremely time consuming. Thus, the final component of the archival project will be in the development of a distributed version of the repository. In this case, the data will likely reside on local sites and tools will be developed to pass analysis results between sites for the purpose of real-time detection and identification.

4.2. Information Sharing Games

An ambitious longer term question will also be addressed: is it possible to scale up our modular global anomaly detection architecture so that it would be attractive to actual Internet operators? To answer this question we will develop a game theoretic and information theoretic framework for studying information sharing among competing autonomous systems in an internetwork. A consortium of ASs may be better able to detect a coordinated attack or anomaly if the ASs share information about their measured internal states, e.g., downed links, switch failures, or excessive throughput delays. Such shared information is beneficial if it can provide early warning of an attack, permitting yet unaffected AS's and users to take appropriate

protective actions, or if it permits the consortium to cooperatively *traceback* the attacker. On the other hand, sharing of information may reveal proprietary information or affect competitive advantage of an individual AS. Thus for any information sharing protocol there exists a tradeoff between insuring information privacy of the individual and ensuring early detection by the group.

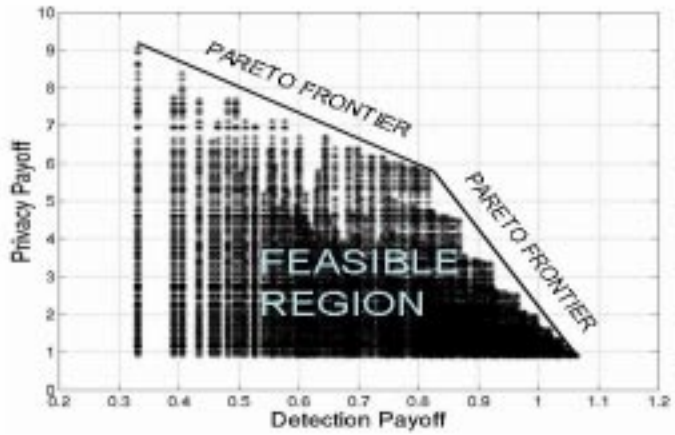

Figure 2: *The feasible region of information sharing payoffs is bounded by the Pareto frontier which delineates the boundary of the feasible payoff region and sets an information sharing benchmark on privacy vs detectability. Asterisks in the feasible region denote the payoff for a set of randomly selected strategies (i.e., information sharing protocols) in a simple linear Gaussian information sharing game for a consortium of 3 AS's undergoing a flooding attack.*

4.2.1. Goals and Challenges in Information Sharing Games: There are many possible ways to design such an information sharing protocol which maintains a certain degree of privacy. For example, each AS could anonymously update a universally known "global" anomaly feature known to all with its own "local" feature. While this simple scheme does not insure total privacy of information, the privacy of local information improves with increasing number of participating AS's. An important question that we will attempt to answer in this project is how to quantify the tradeoff between privacy and early detection for a given class of data sharing protocols. Related questions are: what is the minimum amount of local information that must be revealed in order to improve the global detectability of an attack? Is there an optimal protocol which attains maximum detectability for a given privacy constraint? What is the vulnerability of such a protocol to an attack on an information sharing session?

4.2.2. Research in Information Sharing Games: Quantifying these information tradeoffs is complicated and falls in the domain of cooperative and non-cooperative non-zero sum game theory [49, 96, 48]. When a global anomaly (e.g., a global attack) is suspected, the consortium of all AS's has an incentive to share more local information than any AS might be willing to share under normal conditions. The damage to an AS due to revealing several bits of local information to the consortium can be quantified by the ability of the other AS's (or an attacker) to accurately estimate certain sensitive information states of the AS's network. This can be quantified by the minimum state estimation MSE attainable by the other ASs if they try to snoop on the states of one of the ASs. On the other hand, the value of local information to the consortium of AS's can be quantified by the maximum probability that the group of AS's quickly detect an emerging attack. As the consortium's snooping MSE and anomaly detection probability may be difficult to analyze, when necessary we will adopt an information theoretic approach. Specifically, as in our past work [52, 59, 56, 61], we will adopt the inverse Fisher information as a surrogate for MSE [120, 10] and the Kullback-Leibler information as a surrogate for detection probability [79, 39]. The problem of information sharing can thus be formulated as choosing a set of strategies, i.e., a data sharing protocol, which permits high detection probability but

insures that any attempt by a coalition of AS's to estimate another AS's internal states will result in poor estimates, i.e., high estimation mean square error (MSE). In this information sharing game the players are the AS's, the plays are successive transmissions of information, and the payoff is a vector consisting of the detection probability and estimation MSE. The payoff vectors for a class of protocols lie in the feasible region of the payoff plane, which is bounded by the Pareto frontier (see Figure 2) and can be computed numerically by strategy enumeration, evolutionary algorithms [35] and non-linear programming [158]. Research will include determination of: (i) Pareto frontiers; (ii) associated optimal information sharing strategies; (iii) and countermeasures (e.g. authentication and watermarking games [90, 137, 94, 22]) for reducing vulnerability to attacks on the information sharing session. We will investigate these issues using simple statistical models for nominal (baseline) and anomalous conditions.

5. Distributed Data Analysis

Distributed data analysis is the next crucial step in our framework. The data may undergo some initial processing or statistical summarization, and therefore we will refer to the resulting data or statistics as features. Assume that there are N possible features that may be captured at any point in time and space, and let us arrange these features into an N-dimensional vector $x_{s,t}$, where s refers to the spatial location in the network where the features were collected and t indicates the time at which the features were collected. By 'features' we have in mind any number of measurable quantities such as loss counts, delays, available bandwidth, acknowledgment packets, or numbers of open connections. We use 'space' in the sense of the underlying network topology, which may refer to links or nodes or both. Often a full set of features will not be available at a particular point (due to measurement limitations or privacy concerns), and therefore we express our observations as $y_{s,t} = C_{s,t}x_{s,t}$, where $C_{s,t}$ is a known observability matrix which also may depend on s and t.

5.1. Data Analysis Goals and Challenges: There are two key goals motivating our approach to distributed data analysis:

1. to gain understanding and insight into characteristics, patterns, correlations, and structures in the joint space-time-feature domain and

2. to produce outputs such as summaries, states, or alarms for use at higher levels of comprehensive anomaly/intrusion detection processes.

As examples of the first point, we are interested in detecting and quantifying spatial and temporal correlations in losses and delays, discovering patterns in numbers of connections or certain types of packets related to malicious activity, localizing sources of congestion or failure, and characterizing normal and abnormal traffic flow patterns. Progress on these problems will be useful in their own right, but also will lead naturally to progress on the second point, in the context of our proposed DES framework (described in Section 6).

Analysis of data $\{y_{s,t}\}$ of the sort defined above presents fundamental challenges on several fronts that will be addressed in the proposed work.

Distributed Analysis: The processing and analysis of the array of features $\{x_{s,t}\}$ across many points in time and space is a daunting task. In fact, as the spatial and temporal sampling densities increase it becomes impossible to transmit all features to a central point for processing an analysis in a timely fashion. Therefore, distributed schemes are necessary for on-line, real-time data analysis. Moreover, distributed algorithms take advantage of computing resources throughout the network to perform computations and analysis that would not be possible on at a single, central processor.

Missing Information: At many points the observation $y_{s,t}$ may only convey a limited set of features or may be unavailable all together. Thus, in general we will have an incomplete picture of the time-space-feature domain. Statistical data analysis with missing data is a well known problem, and inferring network states from limited measurements falls into the category of statistical inverse problems. The so-called *network*

tomography problem [19] is a good example of such a case, and it has received increasing attention in the networking, signal processing, and statistics communities [15, 93, 20, 42, 21, 100, 156, 140]. Distributed methods for solving inverse problems and problems of data interpolation or extrapolation in networking is a virtually unexplored research area that is a focus of the proposed work.

Errors and Uncertainties: In addition to missing information, network measurement and monitoring is plagued with a multitude of errors and sources of uncertainty. The distributed nature of data collection and analysis leads to uncertainty at very basic levels such as time synchronization, location (e.g., devices with multiple interfaces), and routing and network "neighborhood" structure. As a consequence the very ideas of absolute time and spatial localization often must be discarded, and the notion of a fixed, static network structure often must be abandoned. These issues confound our usual approaches to correlational and pattern analysis [12, 153, 133], and the research proposed in this proposal will aim to define a new approach to address these unprecedented limitations.

5.2. Proposed Research in Data Analysis: Data analysis can be used to help identify physical mechanisms or explanations for the data, or data analysis can search for patterns or properties in the data themselves without regard to causality. Our proposed work includes both approaches. We will develop data analysis methods that search for understandings and explanations of mechanisms responsible for patterns related to physical aspects of the network, especially those patterns due to locality in space and time. Understanding mechanisms responsible for such patterns can greatly facilitate modular strategies for monitoring networks. Furthermore, we will devise algorithms for identifying patterns in the time-space-feature domain that may be extremely useful for anomaly and intrusion detection, but for which no physical explanation is apparent or sought. The results of both types of data analysis will be fused within the DES framework (described in Section 6).

Data Analysis Using Stochastic Dynamical Models: When spatio-temporal data can be accurately modelled as a continuous Markov random field over space (monitoring site) and time, continuous-variable stochastic models provide a very compact description useful for data analysis. Univariate time series models, such as the fractional autoregressive moving average (FARIMA) model [66, 9, 97], have been previously applied to analyze network traffic statistics such as the sequence of packet-lengths measured at a link [97]. Other useful time series models that have been applied to networks include: ARMA [13], and multifractal (MF) [124, 123, 122, 121, 131, 159, 132]. Such models represent traffic streams as a linear combination of partially observed states, e.g. a vector of the m most recent time samples, which evolve over time. The models are implemented recursively in time and update a predictor of a future unobserved value of the time series based on the sequence of previously observed residual prediction errors. The residual prediction errors can be used for analysis of model fit (model validation), real-time detection of deviations from baseline (anomaly detection), and real-time multiple model selection (anomaly classification). A principal focus will be to extend linear and non-linear multivariate time series models to distributed spatio-temporal stochastic dynamical models for analysis of network traffic. Research issues include: accounting for time synchronization uncertainty; accounting for heavy-tailed traffic distributions); accommodating irregularly sampled traffic at sampling rates that may differ at each site, and distributed implementations that allow for updating the traffic predictor recursively in both space and time.

Distributed, Multiscale Spatio-Temporal Data Analysis: While the task of internetwork monitoring is fundamentally spatio-temporal in nature, vastly more attention has been focused on studying and modeling the temporal aspects of network traffic data. Yet, as has been argued throughout this proposal, successful monitoring and anomoly detection must accurately capture and characterize the spatial aspects as well. And in this the issue of scale plays a critical role. For example, when traffic exhibits unusual characteristics, an immediate question arises as to the size and extent of the region and time period over which the anomaly occurs. Therefore, it is necessary to have scale-sensitive tools for the spatio-temporal analysis of network data. And these tools must in turn be amenable to distributed implementations. There is a wide range

of multiscale data analysis methods that have been developed and applied with great success across the sciences [160, 95, 41, 72, 78, 98, 101, 134]. However, the vast majority are designed for analysis of traditional signal and image data, and are not adapted for use with network topologies. An exception is the recent framework of 'graph wavelets' developed by two of the PIs [27]. This approach extends the concept of a wavelet analysis to arbitrary connected graphs, and was found to be successful for gaining insight into a network's global traffic response to a link failure and for localizing the extent of the failure event within the network. We will work to develop this initial progress in a number of directions, within the context of various continuous-variable stochastic discrete-time models (as described above) for the original traffic data. Proposed extensions include distributed implementations, tree-based hierarchical analogues, and hybrid combinations of these with traditional temporal multiscale methods.

Distributed Pattern Analysis and Learning: The objective of this component of the research is to develop distributed and hierarchical algorithms for finding patterns of baseline and anomalous behavior in time, space and feature domains. We will characterize the trade-offs between information sharing, resource management, and pattern estimation/detection accuracy, and determine fundamental bounds on the achievable performance of distributed pattern analysis systems. Learning approaches to pattern recognition have led to breakthroughs in high dimensional classification problems such as hand written character recognition, genetic sequencing, and image indexing [14, 25, 55]. As examples, tree-based classifiers and kernel-based methods (Support Vector Machines (SVM) and Relevance Vector Machines (RVM) their Bayesian variants [149]) are powerful and computationally efficient nonparametric classifiers. These classifiers do not require specification of a model; using sophisticated complexity-based regularization techniques, trees and SVMs are able to classify complex patterns in high dimensional spaces by learning from training examples alone. Furthermore, trees and SVMs can be used to detect novel or unusual behavior, or other deviations from normal, baseline network characteristics [133]. Recently, we have devised a new approach to constructing tree classifiers that provides concrete bounds on the classification [134] performance (similar bounds are not available for the more well-known CART algorithm [12]). We have also applied CART-like methods to universal prediction and reconstruction of non-linear time series [92] which can be easily adapted to constructing a baseline at the local data collection sites. Developing theory and algorithms for distributed implementations of tree and SVM/RVM classifiers is an open and challenging problem that will be addressed in this work. Our recent work in decentralized, hierarchical methods for detecting non-local phenomena in spatially distributed networks [102] provides a starting point for this aspect of the project.

Distributed Inference with Missing Information and Uncertainty: Estimation and pattern analysis can be extremely complex and computationally challenging when certain data or observations are missing. Computational methods such as the expectation-maximization (EM) algorithm, Markov Chain Monte Carlo (MCMC) methods, and importance sampling can provide computationally efficient approaches to analysis problems involving missing data [40, 88, 104, 146]. Such methods have been widely and successfully applied in a myriad of statistical inverse and missing data problems making them strong candidates for computational tools to use in estimating $\{x_{s,t}\}$ from the observations $\{y_{s,t}\}$, extracting patterns and clustering features, and computing maximum likelihood estimates of summary statistics such as means and covariances. The conventional set-up for computational analysis tools is that all the data and processing are carried out at a central site. We will pursue the development of new theories and computational methods for distributed inference. Our recent work in space-alternating and distributed versions of the EM algorithm [60, 99] provide a starting point for the our investigation into more general strategies for distributed estimation, pattern analysis, and clustering in network data analysis. We will develop theory and algorithms for non-cyclic, non-sequential strategies that allow inferences to be drawn through the limited sharing and communication of information between network elements. These computational methods will also be aimed at coping with other fundamental uncertainties, such as lack of time synchronization between data collection elements and ambiguities in spatial locale and network connectivity.

6. Discrete Event Systems Models for Detection and Diagnosis

The central tenet of our DES approach is that a DES model can provide a compact description of complicated sequences of timed discrete events in the network that are associated with baseline behavior of the network. Such a compact description is possible when the event sequences can be accurately modelled as a semi-Markov chain or when the sequences follow the grammar of a regular language. Similarly to the traffic time series models discussed in the previous section, DES models represent events as combinations of partially observed states which are recursively updated in time. However, DES models differ from traffic time series models in that they can perform efficient real-time inferencing from event sequences derived from a mixture of discrete-valued and continuous-valued data. These data sources are filtered through a *data filtration unit* (DFU) which produces event signals driving the DES. Deterministic DES modelling formalisms include finite-state automata, which have been previously applied to anomaly detection [119, 135] and to LAN intrusion detection [11]. (In [11] for instance, event sequences were derived from ascii `tcpdmp` fields.) Unlike the centralized logical models used in these works, we propose to construct a *distributed hierarchical* DES model which can respond to a *variety of data sources* including, but not restricted to, ascii fields, traffic statistics, and other spatio-temporal features (data summaries) derived from data analysis described in the previous section.

6.1. Goals and Challenges in DES Modelling: The key to our DES approach to internetwork monitoring and anomaly detection is the exploitation of modularity described in Section 3. For concreteness, the discussion below assumes the following three-level hierarchy: local domain (lowest level), AS (middle level), and Network Operations Center (NOC) (highest level). At the local domain level a DFU will process the large quantity of measured data and produce, in real-time, signals that aggregate this information into event signals for subsequent processing by a local DES. These local DES models will have discrete state spaces of logical and quantized variables and event-driven dynamics. The other levels of the hierarchy will also employ DES models. At the AS level, each node of the hierarchy will receive reports from a set of nodes at the local domain level. In turn, sets of nodes at the AS level will report to nodes at the NOC level. In addition, the DES models at any level of the hierarchy will receive other signals from the monitoring of network behavior at that level; see, e.g., [81, 65, 145, 91, 135, 119, 11] and [3, 67, 89] and the references therein for local domain monitoring examples.

Our thesis is that any scalable method for internetwork monitoring benefits from inferencing based on less detailed (i.e., "higher level") dynamical models that properly capture known attacks and other anomalies. The event signals that are reported by the DFU to the DES models might signify deviations from normal traffic patterns and suggest possible hypotheses about causes for these deviations, each accompanied by a confidence index. Hence, we envision the DFU sending event signals of the following form to the DES models: normal behavior; possible attack of type A at this node - level of confidence p ($0 < p \leq 1$ and p quantized appropriately); cancel earlier report of attack of type A; possible attack of type B at this and neighboring nodes - level of confidence p; unclassified anomaly at this node; and so forth. The DES models at the local domain will be constructed in a manner that captures the possible attacks at the node in the form of sequences of *observable* events (i.e., the events reported by the DFU, along with those directly monitored by the DES models), together with sequences of *unobservable* events. Unobservable events are used to model features of the DES operation (e.g., attacks and other anomalies) that are not directly observed or measured but are part of the dynamics of the network; cf. [17]. The DES models will include stochastic information when deemed appropriate. Logical and stochastic automata will be the primary DES modeling formalism used in this regard. At each node, the relevant sequences of observable and unobservable events will be captured in the transition structure of an automaton and lead to appropriately-defined states. Distributed attacks will be modeled by means of *common events* among the respective automata for the local domain nodes affected.

The DES model at each node at the AS level will be a projected/aggregated version of the parallel

composition of the DES models at the local domain level that report to this node. The parallel composition operation is the usual way of synchronizing automata that have events in common. At each of the AS nodes the event signals will be generated by a combination of event sequence information communicated from the DES's at the local sites and perhaps other data collected by the AS, e.g., graph-wavelet coefficients, end-end packet loss or delay statistics obtained by active probing and network tomography. In a similar manner, the NOC level DES model will be a projected/aggregated version of the DES models at the AS level. The purpose of the projection/aggregation is to create models that retain only the information relevant to the joint operation of the given set of nodes and abstract away internal behavior not essential at that particular level of the hierarchy.

Automata and Petri nets (another DES modeling formalism) have been used in prior work on anomaly detection; see, e.g., [81, 65, 145, 91, 135, 119, 11]. Related work of special interest includes asynchronous methods for alarm analysis in the network of France Telecom [8, 111] and receiver and transmitter fault detection in wireless LANs [136, 32]. Our approach is different because: (i) a hierarchy of DES models is used instead of a monolithic model; (ii) the construction of these DES models as well as the detection of network anomalies by each model are based on different information (i.e., information local to the node of the hierarchy); (iii) network attacks and other anomalies are modeled as unobservable events whose occurrence needs to be inferred upon from the sequences of observable events using the dynamical structure of the DES models; and (iv) the DES models are driven by a richer set of observations derived from logical data, ascii data, traffic data, and their statistical summaries.

6.2. Research in DES Modeling:

We propose to investigate the design of computationally- and informationally-efficient modular approaches for monitoring and anomaly detection in internetworks. We will need to answer the following questions: **(Q1)** How is the information contained in the temporal variations of local traffic and packet flows at the local domain processed, i.e., how to specify the DFU? **(Q2)** How and when can monitoring and anomaly detection be done locally, i.e., at each subnetwork separately, without taking into account the subnetwork's coupling with other subnetworks and the information that can be provided by other subnetworks? **(Q3)** When can monitoring and anomaly detection of a subnetwork improve by taking into account its coupling with other subnetworks and by sharing information with other subnetworks? **(Q4)** Suppose the anomaly detection capability at a subnetwork N_i can improve by sharing information with other subnetworks. Which other subnetworks should N_i communicate with and when? What information should be shared in real-time between N_i and other subnetworks? Information sharing must be done in a way that satisfies certain privacy requirements at individual subnetworks. **(Q5)** Given that communication is costly in terms of computation and implementation, what is the minimum information exchange required between N_i and the rest of the network so as to achieve anomaly detection at N_i?

To investigate (Q1) we will initially investigate DFU's that quantize summary statistics, e.g., locally derived likelihood functions and likelihood ratios. For a specified statistical model, these statistics compactly summarize all of the information necessary for inference on the available data, e.g., anomaly detection or classification. For example, this framework can easily be applied to detection using local FARIMA traffic models [66, 97] for which the likelihood ratio reduces to a weighted sum of past residual prediction errors. For anomaly classification a more appropriate approach might be to test between multiple candidate FARIMA models which leads to quantizing the vector of multiple model residual prediction errors. We will also investigate non-parametric approaches such as quantization of graph-wavelet coefficients collected by a node at the intermediate level of the hierarchy. In both the parametric and non-parametric framework a central research issue will be the required resolution of the quantizer in the DFU. Questions that will be addressed include: Should the quantizer be uniform and how many bits should it have? What is the benefit of using a vector quantizer with respect to a scalar product quantizer for the case of vector valued summary statistics? Can methods of distributed source coding [117] or multiple description length source

coding [152] be implemented efficiently on the proposed modular architecture and what are the advantages of these schemes for networked anomaly detectors? We will apply a combination of decentralized decision theory, detection theory, and rate distortion theory [71, 115, 116, 7, 114, 150, 46, 157, 85, 103, 112] including recent extensions [54, 77] to study these questions.

To investigate (Q2), we will begin with the methodology developed in [129, 130, 128, 36, 23]; this methodology has been successfully demonstrated in practical applications: large-scale telecommunication networks [110, 111, 126] and wireless LANs in vehicle platooning [136, 32]. The results to-date in [129, 130, 128, 36, 23] deal with DES modeled by *logical* automata, address the detection and identification of *individual* (unobservable) fault events, and are based on *monolithic* (as opposed to modular) models of the system under consideration. Within the context of network monitoring and anomaly detection, three significant research problems arise: (P1) development of a methodology that considers *stochastic* automata (some preliminary relevant results on diagnosis of stochastic automata are available in [148]); (P2) generalization of the notion of "failure-type labels" (introduced in [129] for tracking unobservable fault events) to *sequences of labels* over space and time that capture partial or complete attack patterns from a database of such patterns; and (P3) development of techniques for exploiting system modularity and avoiding building monolithic models, a task that quickly becomes intractable as the number of subnetworks under consideration grows.

The methodology in [129, 130, 128, 36, 23] relies on the concept of *diagnoser automata* (cf. [129]). Regarding (P1) and (P2) posed above, we conjecture that it will be possible to develop special types of "stochastic diagnoser automata" for the detection of anomalies and answer questions about anomaly detection capabilities (i.e., which anomalies can be detected and which cannot based on the model used and the signals available) using the structure of these stochastic diagnosers. Regarding (P3), we propose to start by building "local" (stochastic) diagnosers at each local domain node based on the DES model at that node, termed *modular diagnosers*, and investigate their anomaly detection capabilities. These may be inadequate since modular diagnosers do not account for other nodes (subnetworks). If they are indeed inadequate, then we must consider the coupling of the node under consideration, say N_i, with other nodes. This coupling, which occurs through the common events in the respective DES model of N_i and of the other nodes, may actually resolve the perceived anomaly detection incapability at N_i. This is the key issue that needs to be investigated. We propose to compose, by parallel composition, suitably "projected" versions of the DES models at N_i and other coupled nodes on the common events they share. By studying the structure of the resulting automaton, we should be able to ascertain the perceived anomaly detection incapability at N_i. If the result of this analysis is that N_i is indeed incapable of achieving the required anomaly detection capabilities, then the modular diagnoser at N_i must be enhanced through real-time information sharing with the other subnetworks. This leads us to questions (Q3)-(Q5) formulated above.

The above questions (Q3)-(Q5) all revolve around the notion of real-time communication among modular diagnosers at different nodes in the hierarchical model that we have adopted. Real-time communication problems in distributed systems are conceptually very difficult. For that reason, to understand them systematically, we will consider separately one-way and two-way communication. One-way communication problems can be formulated as active acquisition of information / querying problems; we will tackle these problems using stochastic control techniques [80]. Two-way communication problems will be formulated as dynamic team problems [64, 147, 164]. We will use the approaches reported in [162, 64, 163, 155, 172, 154, 171, 118, 166, 161, 147, 63, 165, 142], together with heuristics, to determine two-way real-time information sharing strategies that enable modular diagnosers to correctly perform anomaly detection.

7. Education

To enhance the educational impact of this project we will do the following. **(i)** Students doing research on this project will be encouraged to do internships over the summer with our industrial affiliates

Arbor Networks, Internet2, Merit Network, and SLAC (see attached letters). **(ii)** We will organize interdisciplinary seminars (jointly with CS, ECE, and Statistics departments) at each of the three institutions. The seminars will feature expert speakers and bring together students and researchers in networking, signal processing, and statistics. **(iii)** Where feasible, graduate students involved with the project will be co-advised by PIs in different disciplines, and each thesis/dissertation committee will involve at least two of the project PIs (possibly from different institutions). **(iv)** Undergraduate education will be a priority and we will seek supplementary funding for summer undergraduate research projects through the NSF-REU program. **(v)** In addition to active participation in national and international professional meetings, our students will be encouraged to present their work at our annual project review meetings. **(vi)** We will also be engaged in K-12 education. The PIs and students supported by this grant will make presentations to Middle Schools and High Schools (grades 6-12) in local school districts about the importance of internetwork monitoring and anomaly dectection for public safety, and how engineers and scientists tackle such problems. The PI's will also work through two organizations, Camp CAEN and our industry affiliate Merit Networks, Inc., to enhance awareness of computing security for high school students, their teachers, and other high school staff. The College of Engineering at UM organizes Camp CAEN (`http://campcaen.engin.umich.edu/`), a summer camp with a computing focus for students between the ages of 13 and 17. PI's plan to interact with staff at Camp CAEN to develop a program to introduce students to security and monitoring issues. Camp CAEN also provides an all-day, girls-only offering, providing a more supportive environment in which to attract more women to the field. As the Internet service provider for K-12 institutions throughout the state of Michigan, Merit Network, Inc., provides direct access to these institutions. By participating in Merit's K-12 targeted workshops run by their Learning Systems Center to Support Technology in Education, we can advertise educational opportunities, e.g., Camp CAEN or internships, and potentially influence advanced computing curricula at these institutions.

8. Impact of Project

The long term goal of this research is to develop practical and cost-effective internetwork monitoring strategies that can provide the early information that allows a network to make rapid operational changes to maintain performance, avert failure or respond to attack. Our proposed research represents a first step towards this ambitious goal and the intellectual, technical and societal impact of our project will be significant. The short term impact includes: **(i)** a central Web repository for anomaly, attack, and intrusion traces will enhance the network measurement research infrastructure; **(ii)** dissemination of validated software modules for performing offline distributed data analysis on large data sets, including those archived in the repository. Longer term impact includes: **(iii)** development of a modular framework for studying anomalies in large scale networks; **(iv)** development of distributed online monitoring systems which can perform real-time detection and diagnosis of anomalies; **(v)** development of a theory of information sharing which captures both the proprietary concerns and the security concerns of the networking marketplace.

Our project will have the following **broader impact**; **(i)** every effort will be made to involve several female or under-represented minority students, both graduate students (supported totally or in part by the grant) and undergraduate students (supported by a subsequent REU grants together with support from the participating institutions, as appropriate). The PIs have an excellent track record in this regard and collectively they graduated a total of 8 female, 2 Hispanic and 1 African-American Ph.D.s in the last eight years. Moreover, they are currently supervising or co-supervising 7 female Ph.D. students; several of them are prime candidates for receiving support from this grant if funded; **(ii)** the PI's will work to enhance awareness of computing security for middle school and high school students, their teachers, and other high school staff (see Sec. 7); **(iii)** together with their industrial collaborators at Arbor Networks, I2, Merit, and SLAC, the PIs will interact with network managers to transfer technology developed in the project to the operator community; **(iv)** the PIs will continue their active collaborations with international researchers and institutions, including but not restricted to: McGill University (Canada) and INRIA (France)

9. Management Plan

The breakdown of personnel by sub-areas in Table 1 illustrates the balance of our team in the four areas of this project. [1] A feature of our team is that each university has a pair of PI's who cover two of the three research areas, thereby facilitating cross-disciplinary education of graduate students. In addition, all graduate student research assistants (GSRA) will have at least two PI's on their thesis committees (including outside members from other universities) to further the collaborative aims of this project. Furthermore, close contact with our industry affiliates (see below) on real-life network monitoring problems will help maintain practical relevance of the research. The coordination of research, education, and outreach projects which span 3 universities (UM, UW, BU)[2] will be managed according to the structure illustrated in Figure 3 and is explained below.

Electronic Dissemination and Privacy/Proprietary Concerns: A website will be created as an archive for our research reports and articles, sample data traces, interactive software, course materials, and announcements. It is our intent to make much of our collected data and software available to the public, along with terms and conditions of use, on this web site (see Section 4 for more details). Co-PI Paul Barford at UW will maintain this site with help of his students. Paul Barford has extensive experience in Internet data collection and archival (he runs the DOMINO and Surveyor data collection and dissemination projects at UW) and he will deal with the legal and operational privacy/proprietary issues involved in data dissemination.

Outreach to K-12: The PIs and students supported by this grant will be engaged in several types of K-12 outreach activities including: presentations at local schools, interact with UM's Camp CAEN to develop a computing security summer program for high school kids, and interact with Merit Network's "Learning Systems Center to Support Technology in Education" to help K-12 teachers and staff to transition computing and network security to the classroom (See Section 7 for more details).

International Collaboration: Two collaborators, M. Coates and A. Benveniste, have committed to participate in our effort (letters are attached). Mark Coates is Assistant Professor in the Dept. of Electrical Engineering at McGill and has been a close collaborator with Rob Nowak and Alfred Hero on network tomography. His expertise in Monte Carlo Markov Chain (MCMC) optimization and particle filtering will be applied to distributed Bayesian analysis and modeling of multi-site data. Albert Benveniste is Director of Research at the IRISA laboratory at INRIA-Rennes, France. His extensive experience in adapting DES models to fault diagnosis in packet networks will be crucial for our more ambitious effort to apply DES to distributed detection of anomalies in the Internet. Dr Benveniste has been in close technical contact with Professors Lafortune and Teneketzis for the past five years. Funds have been budgeted for international travel to permit face-to-face meetings with these two collaborators.

Industry Affiliates: To enhance the educational and research impact of this project we have invited several representatives from the network operator community, commercial industry, and the networking user community to support this project as "industry affiliates." Representatives from these organizations will meet with us once a year at our annual workshop (see below). Their role will be to help identify ways to improve practical impact of the project, provide guidance on future directions of the research, and provide summer internships for students supported on the grant. We have attached letters of support from the following companies and organizations: Arbor Networks, Internet2 (I2), Merit Network, and Stanford Linear Accelerator Center (SLAC). If the proposal is funded we will invite other organizations to participate.

Annual Review and Workshop: We will organize an annual year-end review to bring our team together to present research results and discuss research objectives. The workshop venue will alternate between

[1] Due to administrative complications connected with the Fullbright Fellowship program, Prof. George Michailides could not be listed among the co-PI's on any submitted budget. He will be returning to the University of Michigan in September (Dept. of Statistics) and will be supported as a co-PI on this grant if funded at requested level.

[2] Prof. Rob Nowak will be moving from Rice to the University of Wisconsin in May 2003 where he will be in close contact with Prof. Paul Barford.

co-PI	Data Collection	Data Analysis	DES Models
P. Barford(UW)	X	X	
M. Crovella(BU)	X	X	
A. Hero(UM)		X	X
E. Kolaczyk(BU)	X	X	
S. Lafortune(UM)	X		X
G. Michailidis(UM)		X	X
R. Nowak(Rice-UW)	X	X	
D. Teneketzis(UM)	X		X
A. Benveniste(INRIA)			X
M. Coates(McGill)		X	

Table 1: Matrix of associations between senior-personnel and collaborators (at bottom in itallics) and sub-areas of this project.

Boston, Ann Arbor, and Madison. Collaborators, industry affiliates, and others will be invited to attend. During this meeting industry affiliates and collaborators will help us evaluate progress of the project based on the following criteria: the effectiveness of collaborations; innovations in theory, algorithms, and data collection; education and outreach; and dissemination (journal and conference publications, software tools, tech-transfer).

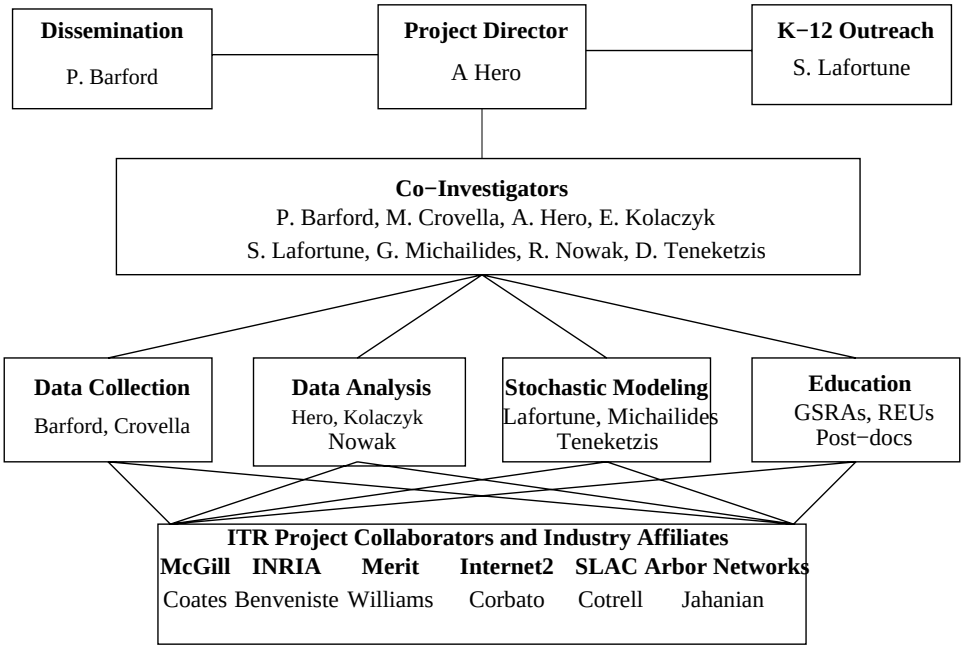

Figure 3: *Management structure for the project. A. Hero will coordinate meetings, workshops, and exchanges between the co-investigators, collaborators, and industry affiliates. He will interface with P. Barford and S. Lafortune who will be the principals responsible for dissemination and outreach to K-12, respectively.*

D Bibliography

References

[1] Abilene. *Internet2 Backbone.* $http://www.internet2.org$, 2003.

[2] J. Almeida, D. Eager, M. Ferris, and M. Vernon. *Provisioning Content Distribution Networks for Streaming Media*, June 2002.

[3] S. Axelsson, "Research in intrusion-detection systems: A survey," Technical Report 98–17, Department of Computer Engineering, Chalmers University of Technology, Goteborg, Sweden, December 1998.

[4] P. Barford, J. Kline, D. Plonka, and A. Rom, "A signal analysis of network traffic anomalies," in *Proceedings of ACM SIGCOMM Internet Measurement Workshop '02*, Marseilles, France, November 2002.

[5] P. Barford and J. Sommers, "A comparison of active and passive methods for measuring packet loss," Technical report, University of Wisconsin-Madison, 2002.

[6] P. Barford, A. Bestavros, J. Byers, and M. Crovella, "On the marginal utility of deploying measurement infrastructure," in *Proceedings of the ACM SIGCOMM Internet Measurement Workshop 2001*, Nov 2001.

[7] G. R. Benitz and J. A. Bucklew, "Asymptotically Optimal Quantizers for Detection of I.I.D. Data," *IEEE Trans. on Inform. Theory*, vol. 35, no. 2, pp. 316–325, Mar. 1989.

[8] A. Benveniste, E. Fabre, C. Jard, and S. Haar, "Diagnosis of asynchronous discrete event systems: A net unfolding approach," in *Proc. 6th International Workshop on Discrete Event Systems (WODES'02)*, M. Silva, A. Giua, and J. Colom, editors, pp. 182–190. IEEE Computer Society, October 2002.

[9] R. J. Beran, *Statistics for Long-Memory Processes*, Chapman & Hall, 1994.

[10] P. J. Bickel and K. A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*, Holden-Day, San Francisco, 1977.

[11] J. W. Branch, A. Bivens, C. Y. Chan, T. K. Lee, and B. K. Szymanski, "Denial of service intrusion detection using time dependent deterministic finite automata," Rensselaer Polytechnic Institute, 2002.

[12] L. Breiman, J. Friedman, R. Olshen, and C. J. Stone, *Classification and Regression Trees*, Wadsworth, Belmont, CA, 1983.

[13] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, Springer-Verlag, New York, 1987.

[14] C. Burges and A. S. (Eds), *Advances in kernel methods - support vector machines*, MIT, Cambridge, 1999.

[15] R. Cáceres, N. Duffield, J. Horowitz, and D. Towsley, "Multicast-based inference of network-internal loss characteristics," *IEEE Trans. Info. Theory*, vol. 45, no. 7, pp. 2462–2480, November 1999.

[16] R. L. Carter and M. E. Crovella, "Measuring bottleneck link speed in packet switched networks," *Performance Evaluation*, vol. 27&28, pp. 297–318, 1996.

[17] C. G. Cassandras and S. Lafortune, *Introduction to Discrete Event Systems*, Kluwer Academic Publishers, 1999.

[18] K. Claffy, *Internet Traffic Characterization*, PhD thesis, University of California, San Diego, 1994.

[19] M. Coates, A. Hero, R. Nowak, and B. Yu, "Internet tomography," *IEEE Signal Processing Magazine*, vol. 19, pp. 47–65, 2002.

[20] M. Coates and R. Nowak, "Network loss inference using unicast end-to-end measurement," in *ITC Seminar on IP Traffic, Measurement and Modelling*, Monterey, CA, Sep. 2000.

[21] M. Coates and R. Nowak, "Sequential Monte Carlo inference of internal delays in nonstationary data networks," *IEEE Transactions on Signal Processing*, vol. 50, pp. 366 –376, 2002.

[22] A. Cohen and A. Lapidoth, "The Gaussian watermarking game Part I, II," submitted to *IEEE Transactions on Information Theory*, preprint 2001.

[23] O. Contant, S. Lafortune, and D. Teneketzis, "Failure diagnosis of discrete event systems: The case of intermittent failures," in *Proc. 41st IEEE Conf. on Decision and Control*, December 2002.

[24] J. Cowie, A. Ogielski, B. Premore, and Y. Yuan. *Global Routing Instabilities during Code Red II and Nimda Worm Propagation*. Available at *http:www.renesys.comprojectsbgp_instability*.

[25] N. Cristiani and J. Shaw-Taylor, *Suport Vector Machines and other kernel-based learning methods*, Cambridge U. Press, 2000.

[26] M. Crovella and A. Bestavros, "Self-similarity in World Wide Web traffic: Evidence and possible causes," *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, pp. 835–846, December 1997.

[27] M. Crovella and E. Kolaczyk, "Graph wavelets for whole-network traffice analysis," in *Proceedings of IEEE INFOCOM 2003*, April 2003.

[28] M. Crovella and E. Kolaczyk, "Graph wavelets for spatial traffic analysis," in *Proceedings of IEEE Infocom*, April 2003.

[29] M. E. Crovella and A. Bestavros, "Self-similarity in World Wide Web traffic: Evidence and possible causes," in *Proceedings of the 1996 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, pp. 160–169, May 1996.

[30] M. E. Crovella and A. Bestavros, "Self-similarity in World Wide Web traffic: Evidence and possible causes," *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, pp. 835–846, December 1997.

[31] M. E. Crovella, M. S. Taqqu, and A. Bestavros, "Heavy-tailed probability distributions in the World Wide Web," in *A Practical Guide To Heavy Tails*, R. J. Adler, R. E. Feldman, and M. S. Taqqu, editors, chapter 1, pp. 3–26, Chapman & Hall, New York, 1998.

[32] H. T. Şimşek, R. Sengupta, S. Yovine, and F. Eskafi, "Fault diagnosis for intra-platoon communication," in *Proc. 38th IEEE Conf. on Decision and Control*, December 1999.

[33] D. Culler. *Towards a Distributed Test-Lab for Planetary-Scale Services* . $http://www.planetlab.org$, 2002.

[34] P. Danzig, J. Mogul, V. Paxson, and M. Schwartz. *The Internet Traffic Archive.* $http://ita.ee.lbl.gov$, 2000.

[35] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*, Wiley, New York, 2001.

[36] R. Debouk, S. Lafortune, and D. Teneketzis, "Coordinated decentralized protocols for failure diagnosis of discrete-event systems," *Discrete Event Dynamic Systems: Theory and Applications*, vol. 10, no. 1/2, pp. 33–86, January 2000.

[37] R. Debouk, S. Lafortune, and D. Teneketzis, "On an optimization problem in sensor selection," *Discrete Event Dynamic Systems: Theory and Applications*, vol. 12, no. 4, pp. 417–445, October 2002.

[38] R. Debouk, S. Lafortune, and D. Teneketzis, "On the effect of communication delays in failure diagnosis of decentralized discrete event systems," *Discrete Event Dynamic Systems: Theory and Applications*, 2003. To appear.

[39] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Springer-Verlag, NY, 1998.

[40] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc.*, vol. 39, pp. 1–38, 1977.

[41] D. Donoho and I. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Statist. Assoc.*, vol. 90, no. 432, pp. 1200–1224, Dec. 1995.

[42] N. Duffield, F. L. Presti, V. Paxson, and D. Towsley, "Inferring link loss using striped unicast probes," in *Proceedings of IEEE INFOCOM 2001*, Anchorage, Alaska, April 2001.

[43] J. A. Fessler and A. O. Hero, "Space alternating generalized expectation- maximization algorithm," *IEEE Transactions on Signal Processing*, vol. 42, no. 10, pp. 2664–2677, October 1994.

[44] J. A. Fessler and A. O. Hero, "Penalized maximum likelihood image reconstruction using space alternating generalized EM algorithms," *IEEE Transactions on Image Processing*, vol. 4, no. 10, , October 1995.

[45] P. Fiorini, L. Lipsky, and M. Crovella, "Consequences of ignoring self-similar data traffic in communications modeling," in *Proceedings of Tenth International Conference on Parallel and Distributed Computing Systems (PDCS-97)*, pp. 322–327, October 1997.

[46] T. J. Flynn and R. M. Gray, "Encoding of Correlated Observations," *IEEE Trans. on Inform. Theory*, vol. 33, no. 6, pp. 773–787, Nov. 1987.

[47] National Laboratory for Applied Network Research. $http://www.nlanr.net$, 2003.

[48] D. Fudenberg and D. Levine, *The theory of learning in games*, MIT Press, Boston, 1998.

[49] D. Fudenberg and J. Tirole, *Game theory*, MIT Press, Boston, 1991.

[50] J. Gast and P. Barford, "Resource deployment based on autonomous system clustering," in *Proceedings of IEEE Globcom '02*, Taipai, Tiawan, October 2002.

[51] S. Gopal and E. D. Kolaczyk, "Understanding the effects of scale in geography: A multiscale analysi s of Massachusetts census data," *Submitted*, 2002.

[52] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. on Inform. Theory*, vol. IT-36, pp. 1285–1301, Nov. 1990.

[53] W. W. C. W. Group. *The Web Characterization Repository*. $http://repository.cs.vt.edu$, 1998.

[54] R. Gupta, *Quantization Strategies for Low-Power Communications*, PhD thesis, Dept. EECS, University of Michigan - Ann Arbor, 2001.

[55] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning : Data Mining, Inference, and Prediction*, Springer Series in Statistics, 2001.

[56] A. O. Hero and R. Delap, "Task specific criteria for adaptive beamsumming with slow fading signals," in *Advances in Spectrum Analysis and Array Processing: Vol. III*, S. Haykin, editor, Prentice Hall, Englewood-Cliffs, NJ, 1995.

[57] A. O. Hero and J. A. Fessler, "Convergence in norm for alternating expectation-maximization (EM) type algorithms," *Statistica Sinica*, vol. 5, no. 1, pp. 41–54, 1995.

[58] A. O. Hero and J. A. Fessler, "A recursive algorithm for computing CR-type bounds on estimator covariance," *IEEE Trans. on Inform. Theory*, vol. 40, pp. 1205–1210, July 1994.

[59] A. O. Hero, J. A. Fessler, and M. Usman, "Exploring estimator bias-variance tradeoffs using the uniform CR bound," *IEEE Trans. on Signal Processing*, vol. 44, pp. 2026–2042, Aug. 1996. `http://www.eecs.umich.edu/~hero/det_est.html`.

[60] A. Hero and J. Fessler, "Convergence in norm for EM-type algorithms," *Statistica Sinica*, vol. 5, pp. 41–54, 1995.

[61] A. Hero, B. Ma, O. Michel, and J. Gorman, "Applications of entropic spanning graphs," *IEEE Signal Processing Magazine*, vol. 19, no. 5, pp. 85–95, Sept. 2002. `http://www.eecs.umich.edu/~hero/imag_proc.html`.

[62] A. Hero, M. Usman, A. Sauve, and J. Fessler, "Recursive algorithms for computing the Cramer-Rao bound," *IEEE Trans. on Signal Processing*, vol. SP-45, no. 3, pp. 803–807, 1997.

[63] Y. C. Ho and T. S. Chang, "Another look at the nonclassical information structure problem," *IEEE Transactions on Automatic Control*, vol. 25, pp. 537–540, 1980.

[64] Y. C. Ho and K. C. Chu, "Information structure in many-person optimization theory," *Automatica*, vol. 10, pp. 341–351, 1974.

[65] K. Ilgun, R. A. Kemmerer, and P. A. Porras, "State transition analysis: A rule-based intrusion detection approach," *Software Engineering*, vol. 21, no. 3, pp. 181–199, 1995.

[66] R. Jana and S. Dey, "Change detection in teletraffic models," *IEEE Trans. on Signal Processing*, Mar. 2000.

[67] A. K. Jones and R. S. Sielken, "Computer system intrusion detection: A survey," Technical report, University of Virginia Computer Science Department, 2000.

[68] J. Ju, E. D. Kolaczyk, and S. Gopal, "Gaussian mixture discriminant analysis and sub-pixel landcover classi fication in remote sensing," in *Proceedings of the* 34^{th} *Interface Symposium*, E. Wegman, editor, 2002.

[69] J. Ju, E. D. Kolaczyk, and S. Gopal, "Gaussian mixture discriminant analysis and sub-pixel landcover charac terization in remote sensing," *Remote Sensing of Environment*, 2003. (In press).

[70] S. Kalidindi and M. Zekauskas, "Surveyor: An infrastructure for internet performance measurements," in *Proceedings of INET '99*, June 1999.

[71] S. A. Kassam, "Optimum Quantization for Signal Detection," *IEEE Trans. on Communications*, vol. COM-25, pp. 479–484, May 1977.

[72] E. Kolaczyk and R. Nowak, "Multiscale likelihood analysis and complexity penalized estimation," *Annals of Statistics* (tentatively accepted for publication). Also available at `www.ece.rice.edu/~nowak/pubs.html`, 2002.

[73] E. D. Kolaczyk, "On the use of prior and posterior information in the sub-pixel problem," *IEEE Transactions on Geoscience and Remote Sensing (Letters)*, 2003. (Under review).

[74] E. D. Kolaczyk and R. D. Nowak, "Multiscale likelihood analysis and complexity penalized estimation," *Annals of Statistics*, 2002. (tentatively accepted).

[75] E. D. Kolaczyk and R. D. Nowak, "Multiscale statistical models," in *MSRI Conference on Nonlinear Estimation and Classification*, D. Denison et al., editors, Springer Verlag, New York, New York, 2002.

[76] E. Kolaczyk and H. Huang, "Multiscale statistical models for hierarchical aggregation," *Geographical Analysis*, vol. 33, pp. 95–118, 2001.

[77] T. Kragh and A. O. Hero, "Emission tomography from compressed data," in *Proc. of Asilomar Conference*, p. to appear, Monterey, CA, November 2002.

[78] H. Krim and I. Schick, "Minmax description length for signal denoising and optimal representation," *IEEE Trans. on Information Theory*, vol. 45, no. 3, , April, 1999.

[79] S. Kullback, *Information Theory and Statistics*, Dover, 1978.

[80] P. R. Kumar and P. Varaiya, *Stochastic Systems. Estimation, Identification, and Adaptive Control*, Prentice-Hall, 1986.

[81] S. Kumar and E. H. Spafford, "A Pattern Matching Model for Misuse Intrusion Detection," in *Proceedings of the 17th National Computer Security Conference*, pp. 11–21, 1994.

[82] S. Lafortune, D. Teneketzis, M. Sampath, R. Sengupta, and K. Sinnamohideen, "Failure diagnosis of dynamic systems: An approach based on discrete event systems," in *Proc. 2001 American Control Conf.*, pp. 2058–2071, June 2001.

[83] A. Lakhina, J. Byers, M. Crovella, and P. Xie, "Sampling biases in IP topology measurements," in *Proceedings of IEEE Infocom*, April 2003.

[84] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Transactions on Networking*, pp. 2:1–15, 1994.

[85] M. Longo, T. D. Lookabaugh, and R. M. Gray, "Quantization for Decentralized Hypothesis Testing under Communication Constraints," *IEEE Trans. on Inform. Theory*, vol. 36, pp. 241–255, March 1990.

[86] M. M. Louie and E. D. Kolaczyk, "Multiscale spatial process models," *Journal of Multivariate Analysis*, 2002. (Under review).

[87] A. Mahanti, D. Eager, M. Vernon, and D. Sundaram-Stukel. *Scalable On-Demand Media Streaming with Packet Loss Recovery*, August 2001.

[88] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, Wiley, New York, 1997.

[89] L. Mé and C. Michel, "Intrusion detection: A bibliography," Technical Report SSIR-2001-01, Supelec, Rennes, France, September 2001.

[90] N. Merhav, "On random coding error exponents of watermarking systems," IEEE Trans. on Inform. Theory, Vol IT-46, No 2, 420430, Mar 2000.

[91] C. Michael and A. Ghosh, "Using finite automata to mine execution data for intrusion detection: A preliminary report," *Lecture Notes in Computer Science*, vol. 1907, , 2000.

[92] O. Michel, A. Hero, and A.-E. Badel, "Tree structured non-linear signal modeling and prediction," *IEEE Trans. on Signal Processing*, vol. SP-47, no. 11, pp. 3027–3041, Nov. 1999.

[93] *Multicast-based inference of Network-internal Characteristics (MINC)*. http://gaia.cs.umass.edu/minc.

[94] P. Moulin and J. A. O'Sullivan, "Information-theretic analysis of watermarking," in *Proc. IEEE International Conference on Acustics, Speech and Signal Processing*, vol. 6, pp. 36303633, 2000.

[95] P. Moulin and J. Liu, "Statistical imaging and complexity regularization," *IEEE Transactions on Information Theory*, vol. 46, no. 5, pp. 1762 –1777, 2000.

[96] R. Myerson, *Game theory: the analysis of conflict*, Harvard Univ. Press, Cambridge, 1991.

[97] K. Nagarajan and T. Zhou, "A new resource allocation scheme for VBR video sources," in *Proc. Asilomar Conf. on Signals, Systems, and Computers (ASILOMAR)*, Oct. 2000.

[98] R. Nowak, "Multiscale hidden Markov models for Bayesian image analysis," in *Bayesian Inference in Wavelet Based Models*, pp. 243–266. Springer-Verlag, 1999. Editors P. Müller and B. Vidakovic.

[99] R. Nowak, "Distributed EM algorithms for density estimation and clustering in sensor networks," to appear in *IEEE Trans. Sig. Proc.*, 2003.

[100] R. Nowak and M. Coates, "Unicast network tomography using the em algorithm," submitted to *IEEE Transactions on Information Theory*, 2001.

[101] R. Nowak and E. Kolaczyk, "A Bayesian multiscale framework for Poisson inverse problems," *IEEE Transactions on Information Theory*, 2000.

[102] R. Nowak and U. Mitra, "Boundary estimation in sensor networks: Theory and methods," in *Proc. 2nd International Workshop on Information Processing in Sensor Networks*, Palo Alto, CA, April 2003.

[103] K. Oehler and R. M. Gray, "Combining Image Compression and Classification using Vector Quantization," *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol. 17, no. 5, pp. 461–473, May 1995.

[104] F. O'Sullivan, "A statistical perspective on ill-posed inverse problems," *Statistical Science.*, vol. 1, no. 4, pp. 502–527, 1986.

[105] K. Park, G. T. Kim, and M. E. Crovella, "On the relationship between file sizes, transport protocols, and self-similar network traffic," in *Proceedings of the Fourth International Conference on Network Protocols (ICNP'96)*, pp. 171–180, October 1996.

[106] K. Park, G. T. Kim, and M. E. Crovella, "The protocol stack and its modulation effect on self-similar traffic," in *Self-Similar Network Traffic and Performance Evaluation*, K. Park and W. Willinger, editors, Wiley / Wiley Interscience, New York, 1999.

[107] K. Park, G. Kim, and M. E. Crovella, "On the effect of traffic self-similarity on network performance," in *Proceedings of SPIE International Conference on Performance and Control of Network Systems*, November 1997.

[108] V. Paxson. *Some Not So Pretty Admissions about Dealing with Internet Measurements*. Invited talk: Workshop on Network Related Data Management, 2001.

[109] V. Paxson and S. Floyd, "Wide-area traffic: The failure of poisson modeling," *IEEE/ACM Transactions on Networking*, vol. 3(3), pp. 226–244, June 1995.

[110] Y. Pencolé, "Decentralized diagnoser approach: Application to telecommunication neworks," in *Proc. DX'00: Eleventh International Workshop on Principles of Diagnosis*, A. Darwiche and G. Provan, editors, pp. 185–192, June 2000.

[111] Y. Pencolé, M.-O. Cordier, and L. Rozé, "A decentralized model-based diagnostic tool for complex systems," in *Proc. 13th IEEE Int. Conf. on Tools with Artif. Intel. (IC-TAI'01)*, pp. 95–102, 2001.

[112] K. Perlmutter, S. Perlmutter, R. Gray, R. Olshen, and K. Oehler, "Bayes risk weighted vector quantization with posterior estimation for image compression and classification," *IEEE Trans. on Image Processing*, vol. IP-5, no. 2, pp. 347–360, 1996.

[113] D. Plonka, "Flowscan: A network traffic flow reporting and visualization tool," in *Proceedings of the USENIX Fourteenth System Administration Conference LISA XIV*, New Orleans, LA, December 2000.

[114] B. Picinbono and P. Duvaut, "Optimum Quantization for Detection," *IEEE Trans. on Communications*, vol. 36, no. 11, pp. 1254–1258, Nov. 1988.

[115] H. V. Poor and J. B. Thomas, "Applications of Ali-Silvey Distance Measures in the Design of Generalized Quantizers," *IEEE Trans. on Communications*, vol. COM-25, pp. 893–900, Sep. 1977.

[116] H. V. Poor, "Fine Quantization in Signal Detection and Estimation – Part 1," *IEEE Trans. on Inform. Theory*, vol. 34, pp. 960–972, Sep. 1988.

[117] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," in *Proceedings of the IEEE Data Compression Conference*, pp. 193 – 262, March 1999.

[118] R. Radner, "Team," in *Decision and Organization*, C. B. McGuire and R. Radner, editors. U. of Minnesota Press, 1986.

[119] V. Ramezani, S.-A. Yang, and J. Baras, "Finite automata models for anomaly detection," Technical Report TR 2002-42, University of Maryland, October 2002.

[120] C. R. Rao, *Linear Statistical Inference and Its Applications*, Wiley, New York, 1973.

[121] V. Ribeiro, R. Riedi, M. Coates, and R. G. Baraniuk, "Multifractal cross-traffic estimation from end-to-end measurements," *Proceedings ITC Specialist Seminar on IP Traffic Measurement, Modeling and Management, Monterey, CA*, Sept. 2000.

[122] V. Ribeiro, R. Riedi, M. S. Crouse, and R. G. Baraniuk, "Multiscale queuing analysis of long-range-dependent network traffic," *Proceedings of IEEE INFOCOM 2000, Tel Aviv, Israel*, March 2000.

[123] V. Ribeiro, R. Riedi, M. S. Crouse, and R. G. Baraniuk, "Simulation of non-Gaussian long-range-dependent traffic using wavelets," *Proc. SigMetrics*, pp. 1–12, May 1999.

[124] R. Riedi, M. S. Crouse, V. Ribeiro, and R. G. Baraniuk, "A multifractal wavelet model with application to TCP network traffic," *IEEE Trans. Info. Theory, Special issue on multiscale statistical signal analysis and its applications*, vol. 45, pp. 992–1018, April 1999.

[125] Route Views. *University of Oregon.* http://www.antc.uoregon.edu/routeviews, 2003.

[126] L. Rozé and M.-O. Cordier, "Diagnosis discrete-event systems: Extending the diagnoser approach to deal with telecommunication networks," *Discrete Event Dynamic Systems: Theory and Applications*, vol. 12, pp. 43–81, 2002.

[127] C. Samios and M. Vernon, "Modeling the throughput of TCP Vegas," in *Proceedings of ACM SIGMETRICS (to appear)*, San Diego, CA, June 2003.

[128] M. Sampath, S. Lafortune, and D. Teneketzis, "Active diagnosis of discrete event systems," *IEEE Trans. Automatic Control*, vol. 43, no. 7, pp. 908–929, July 1998.

[129] M. Sampath, R. Sengupta, S. Lafortune, K. Sinnamohideen, and D. Teneketzis, "Diagnosability of discrete event systems," *IEEE Trans. Automatic Control*, vol. 40, no. 9, pp. 1555–1575, September 1995.

[130] M. Sampath, R. Sengupta, S. Lafortune, K. Sinnamohideen, and D. Teneketzis, "Failure diagnosis using discrete event models," *IEEE Trans. Control Systems Technology*, vol. 4, no. 2, pp. 105–124, March 1996.

[131] S. Sarvotham, R. Riedi, and R. Baraniuk, "Connection-level analysis and modeling of network traffic," *Proceedings IEEE/ACM SIGCOMM Internet Measurement Workshop, San Francisco*, Nov 2001.

[132] S. Sarvotham, X. Wang, R. Riedi, and R. Baraniuk, "Additive and multiplicative mixture trees for network traffic modeling," *Proceedings ICASSP Orlando, FL*, May 2002.

[133] B. Scholkopf and A. Smola, *Learning with Kernels*, MIT Press, Cambridge, MA, 2002.

[134] C. Scott and R. Nowak, "Dyadic classification trees via structural risk minimization," in *Proc. Neural Information Processing Systems (NIPS)*, Vancouver, CA, Dec. 2002.

[135] R. Sekar, M. Bendre, D. Dhurjati, and P. Bollineni, "A fast automaton-based method for detecting anomalous program behaviors," in *IEEE Symposium on Security and Privacy*, pp. 144–155, 2001.

[136] R. Sengupta, "Diagnosis and communication in distributed systems," in *Proc. of the 1998 International Workshop on Discrete Event Systems (WODES'98)*, pp. 144–151. IEE, August 1998.

[137] A. Sequiera, D. Kundar and E. Rogers, "Communication and Information Theory in Watermarking: A Survey," *Multimedia Systems and Applications IV*, A. G. Tescher, B. Vasudev, and V. M. Bove, eds., Proc. SPIE (vol. 4518), pp. 216-227, Denver, Colorado, August 2001.

[138] M.-F. Shih and A. O. Hero, "Unicast inference of network link delay distributions from edge measurements," in *Proc. IEEE Int. Conf. Acoust., Speech, and Sig. Proc.*, Salt Lake City, UT, May 2001. `http://www.eecs.umich.edu/~hero/comm.html`.

[139] M.-F. Shih and A. O. Hero, "Unicast-based inference of network link delay distributions using finite-mixture models," in *Proc. IEEE Int. Conf. Acoust., Speech, and Sig. Proc.*, Orlando, FA, May 2002. `http://www.eecs.umich.edu/~hero/comm.html`.

[140] M.-F. Shih and A. O. Hero, "Unicast-based inference of network link delay distributions using mixed finite mixture models," *IEEE Trans. on Signal Processing*, To appear Sept. 2003. `http://www.eecs.umich.edu/~hero/comm.html`.

[141] *The Skitter Project*. $http://www.caida.org/tools/measurement/skitter$, 2003.

[142] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Information Theory*, vol. IT-19, pp. 471–480, 1973.

[143] W. Stallings, *SNMP, SNMPv2, SNMPv3, and RMON 1 and 2*, Addison Wesley Longman, Inc., third edition, 1999.

[144] S. Staniford, V. Paxson, and N. Weaver, "How to Own the Internet in Your Spare Time," in *Proceedings of the 11th USENIX Security Symposium*, 2002.

[145] S. Staniford-Chen, S. Cheung, R. Crawford, M. Dilger, J. Frank, J. Hoagland, K. Levitt, C. Wee, R. Yip, and D. Zerkle, "GrIDS – A graph-based intrusion detection system for large networks," in *Proceedings of the 19th National Information Systems Security Conference*, volume 1, pp. 361–370, October 1996.

[146] M. Tanner, *Tools for Statistical Inference*, Springer, New York, 1996.

[147] D. Teneketzis, "Information structures and nonsequential stochastic control," *CWI Quart*, vol. 9, no. 3, pp. 241–260, 1996. Special Issue on Systems and Control.

[148] D. P. L. Thorsley and D. Teneketzis, "Failure diagnosis of stochastic automata," Technical Report CGR-03-05, University of Michigan, January 2003.

[149] M. E. Tipping, "The relevance vector machine," in *In Advances in Neural Information Processing Systems*, Morgan Kaufmann, San Mateo, CA, 2000.

[150] J. N. Tsitsiklis, "Extremal Properties of Likelihood Ratio Quantizers," *IEEE Trans. on Communications*, vol. 41, no. 4, pp. 550–558, Apr. 1993.

[151] J. Ullrich. *DSHIELD*. $http://www.dshield.org$, 2000.

[152] V. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. on Inform. Theory*, pp. 821–824, May 1993.

[153] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, 1995.

[154] P. Varaiya and J. Walrand, "On delayed sharing patterns," *IEEE Transactions on Automatic Control*, vol. 23, no. 3, pp. 443–445, June 1978.

[155] P. Varaiya and J. Walrand, "A minimum principle for decentralized stochastic control," in *Dynamic Optimization and Mathematical Economics*, P. T. Lin, editor, pp. 253–266. Plenum Press, 1980.

[156] Y. Vardi, "Network tomography: estimating source-destination traffic intensities from link data," *J. Amer. Stat. Assoc.*, pp. 365–377, 1996.

[157] R. Viswanathan and A. Ansari, "Distributed detection of a signal in generalized Gaussian noise," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-37, no. 5, pp. 775, 1989.

[158] N. N. Vorob'ev, *Game theory: lectures for economists and systems scientists*, Springer, New York, 1977.

[159] X. Wang, S. Sarvotham, R. Riedi, and R. Baraniuk, "Connection-level modeling of network traffic," *Proceedings DIMACS Workshop on Internet and WWW Measurement, Mapping and Modeling, Rutgers, NJ*, Feb 2002.

[160] A. S. Willsky, "Multiresolution markov models for signal and image processing," *Proceedings of the IEEE*, vol. 90, pp. 1396 –1458, 2002.

[161] H. S. Witsenhausen, "A counter example in stochastic control," *SIAM J. Control*, vol. 6, pp. 131–147, 1968.

[162] H. S. Witsenhausen, "Separation of estimation and control for discrete time systems," *Proc. IEEE*, vol. 59, no. 11, pp. 1557–1566, 1971.

[163] H. S. Witsenhausen, "A standard form for sequential stochastic control," *Mathematical Systems Theory*, vol. 7, no. 1, pp. 5–11, 1973.

[164] H. S. Witsenhausen, *Lecture Notes in Economics and Mathematical Systems*, volume 107, chapter The Intrinsic Model for Discrete Stochastic Control: Some Open Problems, pp. 322–335, Springer-Verlag, Berlin, 1975.

[165] H. S. Witsenhausen, "A simple bilinear optimization problem," *Systems Control Letters*, vol. 8, pp. 1–4, 1986.

[166] H. S. Witsenhausen, "Equivalent stochastic control problems," *Math. Contr. Signals Systems*, vol. 1, pp. 3–11, 1988.

[167] J. Xu, J. Fan, M. Ammar, and S. Moon, "On the design and performance of prefix preserving ip traffic trace anonymization," in *Proceedings of ACM SIGCOMM Internet Measurement Workshop '01*, San Francisco, CA, November 2001.

[168] V. Yegneswaran, P. Barford, and S. Jha, "Global intrusion detection in the domino overlay system," Technical report, University of Wisconsin-Madison, February 2003.

[169] T.-S. Yoo and S. Lafortune, "NP-completeness of sensor selection problems arising in partially-observed discrete event systems," *IEEE Trans. Automatic Control*, vol. 47, no. 9, pp. 1495–1499, September 2002.

[170] T.-S. Yoo and S. Lafortune, "Polynomial-time verification of diagnosability of partially-observed discrete event systems," *IEEE Trans. Automatic Control*, vol. 47, no. 9, pp. 1491–1495, September 2002.

[171] T. Yoshikawa, "Decomposition of dynamic team decision problems," *IEEE Transactions on Automatic Control*, vol. 23, no. 4, pp. 627–632, August 1978.

[172] T. Yoshikawa and H. Kobayashi, "Separation of estimation and control for decentralized stochastic control systems," *Automatica*, vol. 14, pp. 623–628, 1978.