PROJECT DESCRIPTION

1 Introduction

To gain back the trustworthiness of Internet infrastructures and network-centric computer systems, it is critical to improve the *measurable* performance of network systems under cyber atatcks and threats. This project focuses on the design, development, and evaluation of a novel enterprise cyber-defense system, integating flexible intrusion detection, information fusion, QoS-adaptive resource management, and proxy-based intrusion tolerance components.

1.1 motivations

The proliferation of Internet applications and network-centric mission-critical services is bringing network and system security issues to the fore. The past few years have seen significant increase in cyber attacks on the Internet, resulting in degraded confidence and trusts in the use of the Internet and computer systems. The cyber attacks, including email virus, worms, and DDoS, are getting more sophisticated, spreading quicker, and causing more damage. Attacks originally exploited the weakness of the individual protocols and operating systems but now have started to attack the basic infrastructure of the Internet. To gain back the trustworthiness of Internet infrastructures and computer systems, there is an urgent need to enhance the effectiveness of the cyber defense and critical to improve the measurable performance of network systems when under attacks.

Improving the measurable performance against cyber threats requires a multi-faceted research program that include:

- improvement of detection of coordinated attacks, while limiting the cost and impact of the detection process.
- improvement of core network and system Quality-of-Service (QoS) support models for intrusion mitigation under uncertain threats.
- development of novel intrusion tolerance approaches to reduce the impact of the inevitable attacks.
- development of evaluation tools to allow us to have at least statistical confidence in the experimental results.

There are, of course, many other research components that will also improve performance, but the four items above need to be a part of any serious solution to this growing problem. They are the focus of this proposal.

As we more and more rely on the intrusion detection systems (IDS), the IDS systems also give us more and more false alarms. This is due to the lack of flexibility under new cyber attacks and uncertain threats. Some cyber attacks are initially not obvious and disruptive. And, it is often too late when they are known to be malicious attacks. Somce *malware* (malicious programs and code propagating on the Internet) [16, 32] can spread at exponential rates. They can quickyly propagate through the network, infecting many machines before the severity of the situation is recognized. Most today's DDoS attacks aim to completely disable the victim system's service to its clients by consuming its available resources, which are called disruptive attacks. Degrading DDoS attacks (3DoS) are new and emerging. The goal of degrading attacks

is to increasingly or periodically consume portions of a victim system's resources so as to result in denial of service to legitimate clients during high load periods. Some legitimate clients may also leave the victim system due to the experienced poor QoS. 3DoS attacks can remain undetected for a long time period since they do not lead to total service disruption and therefore it is difficult to identify the attackers. Thus, IDS systems have to be flexible under uncertain cyber threats. The existing yes/no intrusion detection models are not sufficient.

A flexible intrusion detection model needs the support of QoS-adaptive resource management in network routers and end-point systems. On one hand, QoS is the target of cyber attacks. Cyber attacks, such as DDoS, aim to reduce QoS level provided by networks and systems and experienced by users; in the worst case, no service at all. On the other hand, QoS can be used as a means against cyber attacks. Under uncertain attack senarios, a router or an end-point system can handle incoming traffic differently according to the confidence levels about observed traffic behaviors provided by the flexible IDS systems. For example, the confidence level can be utilized to limit the propagation rate of a potentially malicious and susceptive traffic. Thus, our proposal is to make the performance of networks and systems configurable and controllable by themselves, instead of by parameters and behvaiors of uncertain attacks.

We further propose to build a bridge from distributed IDS systems to the QoS-adaptive resource management component in network systems. Today's enterprise intrusion detection systems generate large volumes of data from distributed intrusion detection and traffic monitoring devices. It often takes a long time to analyze the intrusion data and this results in slow response time. Efficient information fusion techniques can help correlate distributed intrusion and traffic data and pass along urgent alerts. This enables early warning and intrusion handling. Furthermore, based on the brief network techniques, information fusion component will get correlated intrusion data from IDS systems and generate control parameters to the QoS-adatptive resource management component.

proxy-based Multi-path routing as intrusion tolerance...waiting for Chow's input.

The above motivates particular techniques to detect and mitigate cyber attacks and threats. But equally important in improving the predictability, and trust in cyber security, is the design and implementation of evaluation tools to allow the computation of (statistical) confidence intervals. It is not sufficient to know the IDS detected, say 65% of the attacks in a particular test, we need to know the expected variation in that detection rate over some range of experiments. In addition, as we move to adaptive detection and response techniques, we need to develop tools that allow for efficient generation of online evaluations as well as offline datasets for other evaluations.

1.2 Objectives

The goal of this project is to design, develop, and evaluate an integrated enterprise cyber-defense system for improving measurable performance of network systems. Figure 1 illustrates the architecture of the integrated system.

1. Develop a cooperative intrusion detection and response (IDR) system, which gathers and analyzes the traffic patterns and exchanges alert information with other intrusion defense systems. It delivers the detection results to the information fusion component for further correlation. With the support of QoS-adaptive resource management capability of intrusion mitigation component, the outputs of the

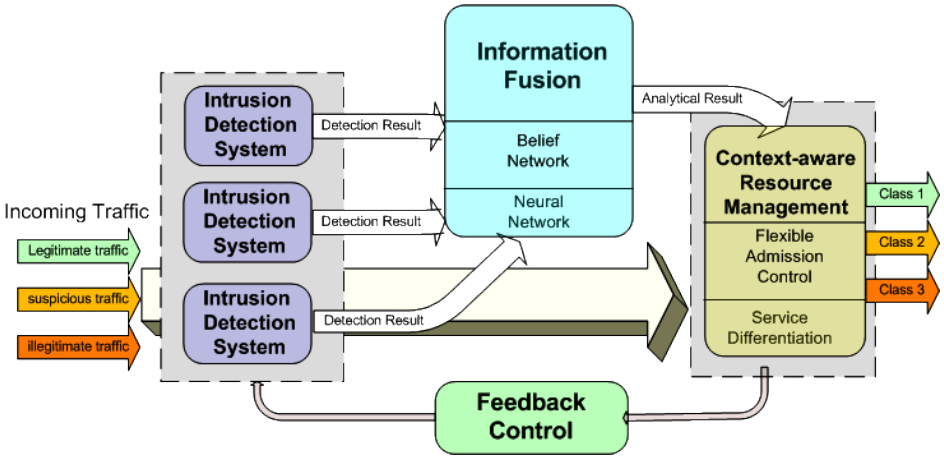

Figure 1: The architecture of the Secure Collective Defense System.

IDR system can be flexible.

2. Develop an information fusion component. **Chow's input**.

3. Develop an integrated admission control with QoS-adaptive resource management component on network systems to extend the essential yes/no model in current IDS systems.

4. Develop proxy-based multi-path routing mechanisms for intrusion tolerance.

2 Background & Related Work

2.1 Intrusion Detection and Response

Recent intrusion detection research has been heading towards a distributed framework on monitors that do local detection and provide information for global detection of intrusions. These includes DIDS [72], GrIDS [24], EMERALD [62] and AAFID [74]. They rely on some predefined hierarchical organization and most of them perform centralized intrusion analysis. Gopalakrishna and Spafford [Gopa01] present a framework for doing distributed intrusion detection with no centralized analysis component. Ning et al [58] presents a decentralized method for autonomous but cooperative component systems to detect distributed attacks specified by signatures.

In [81] a system architecture and mechanisms for protecting mobile ad hoc networks is proposed. Experiment results demonstrates that an anomaly detection approach works well on different mobile ad hoc network with route logic compromise and traffic pattern distortion. It will be interested to see how their framework can be applied in SCOLD systems.

Julisch propose a novel alarm clustering method [37] that remove redundant alarms and support the human analysis in identifying root causes. Experiments show significant reduction in alert load. Effective generalization hierarchies for IP addresses, ports, and time duration are used in the alarm clustering.

2.2 Information Fusion and Correlation

Chow's input

2.3 QoS-adaptive Resource Management

There are recent efforts on providing QoS differentiation from server side. For QoS differentiation provisioning on multimedia servers, the efforts were mostly based on application-level quality adaptation techniques [22, 85]. In [85], we proposed a bandwidth allocation strategy for providing proportional streaming bit rates from a streaming server to clients. The research work is enabled by the advance of real-time video adaptation technology. Multimedia connections impose very different workload characteristics on servers compared to those imposed by conventional Web servers. Techniques developed in the multimedia community are not applicable in our QoS differentiation for Web services context.

On Web servers, response time and slowdown are two fundamental performance metrics of responsiveness. Existing responsive time differentiation strategies are mostly based on admission control, priority scheduling, and content adaptation [1, 14, 25, 43, 40, 83, 84, 86]. In [25], the authors addressed strict priority scheduling strategies for controlling CPU utilization on Internet servers. The results showed that responsive time differentiation can be achieved but the quality spacings among different classes cannot be quantitatively controlled. Time-dependent priority scheduling has been used in achieving proportional delay differentiation in packet networks. It adjusts the priority of a backlogged class according to experienced delays of backlogged packets; see WTP [29] and adaptive WTP [44] for representatives. The algorithms can be tailored in achieving queueing-delay differentiation at the service side [25, 43]. However, they are not applicable for response time differentiation because the response time is not only dependent on a job's queueing delay but also on its service time, which varies significantly depending on the requested services.

Slowdown is the ratio of a request's queueing delay to its service time. Both queueing delay and response time are not suitable to compare requests that have very different resource demands. Actually, clients are likely to anticipate short delays for "small" requests, and are willing to tolerate long delays for "large" requests. Thus, it is desirable that a request's delay be proportional to its processing requirement. A high slowdown can also indicate that the system is heavily loaded. There are few efforts on slowdown differentiation [83, 84, 86].

2.4 Proxy-based Multiple Path Routing

Chen [23] proposed the design of a multipath transport protocol called MPTCP (Multiple Path TCP) that opens multiple TCP connections over different paths and multiplexes data among the paths. Simulation results showed effective use of the available bandwidth on multiple paths even under heavy network utilization levels. However, no actual implementation was carried out. Note that reliable data transfer with multiple paths can be realized on top of TCP as MPTCP, or between TCP and IP such as that proposed by ATCP. It is important to analyze the design trade-offs and compare the performance of these two different approaches.

Transport connections set up in wireless *ad hoc* networks are plagued by problems such as high bit error rates, frequent route changes, and network partitions. If we run transmission control protocol (TCP) over such connections, the throughput of the connection is generally extremely poor because TCP treats lost or delayed acknowledgments as congestion. Several papers [11, 12, 17, 79] have proposed methods for improving TCP performance in *cellular* networks where the last link is the only wireless link in the system. Typically, the solution used in these various papers is to *split* the connection in two at the base

station. The base station then retransmits packets to the mobile node in order to prevent the TCP sender located in the wire line network from invoking congestion control. This approach makes sense because the base station typically knows the state of the wireless link and can make intelligent decisions regarding the state of the TCP connection. In an ad hoc network, on the other hand, the TCP connection traverses multiple wireless links. Thus, solutions based on using the base station to "fix things" do not work well. Liu and Singh [36] proposed an approach called ATCP, where a thin layer is implemented between the Internet protocol and standard TCP that corrects these problems and maintains high end-to-end TCP throughput. It uses explicit link failure notification (ELFN) to improve TCP performance. Here, the sender is notified that a link has failed whereupon it disables its retransmission timer and enters a standby mode. In the standby mode, the TCP sender periodically sends a packet in its congestion window to the destination. When an ACK is received, the TCP protocol leaves the standby mode, restores its retransmission timers, and resumes transmission as normal. The research results have demonstrated that the awareness of location and the underlying network can help improve TCP performance.

In [10], a diversity coding technique was shown to recover information in $N + M$ total blocks through linear transformation, when only M or fewer blocks were lost. In [78], a multiple path routing scheme was proposed for mobile *ad hoc* networks based on diversity coding, where the data load was distributed over multiple paths in order to minimize the packet drop rate, thus achieving load balancing and improving end-to-end delay.

3 Proposed Work

3.1 Cooperative Intrusion Detection and Response

Existing intrusion detection systems are plagued by too many false positives. Techniques are need to clustering the reports, remove the redundant alerts generated by the same root cause. Allowing distributed coordination and direct communications among the IDR devices will help track down the intrusion sources, push back intrusion traffic, detect compromised or malfunction nodes, and provide alternate routes for intrusion tolerance. It will be also of interest to investigate how the collection of proxy servers and the availability of the multiple path indirect routes can be used to improve the security of the network system.

Preliminary Results: In [20], we have developed an Autonomous Anti-DDoS system, called A2D2, where an enhanced SNORT IDS with subnet spoofing plug-in is integrated with a multiple level adaptive rate limiting firewall. Alerts generated by the enhanced SNORT system automatically trigger the insertion of the firewall rules. Users of A2D2 can specified the multiple level of rate limiting. The system keeps history records and adaptively block potential intrusion or put them in queues with restrict packet rates. Preliminary experiment results shows the A2D2 can tolerate various DDoS attacks. A subset of Intrusion Detection and Isolation Protocol [57] were developed and being used in cooperative intrusion push back experiments.

In [26], we have investigated how to deploy firewalls for protecting mobile ad hoc networks. We have developed a PEAP module for the freeRadius server and analyzed the performance of PEAP and TTLS protocol performance for supporting secure wireless access.

Secure AODV protocol was proposed by Zapata et al in [80] to enhance the security of routing update but not real implementation was mentioned. Karlof and Wagner present attacks and countermeasures for secure routing in wireless sensor networks [38]. Perrig et al presents the SPIN protocol for wireless sensor networks [59]. Carmen, Kruus, and Matt analyzed the constraints and approaches for distributed sensor network security [18].

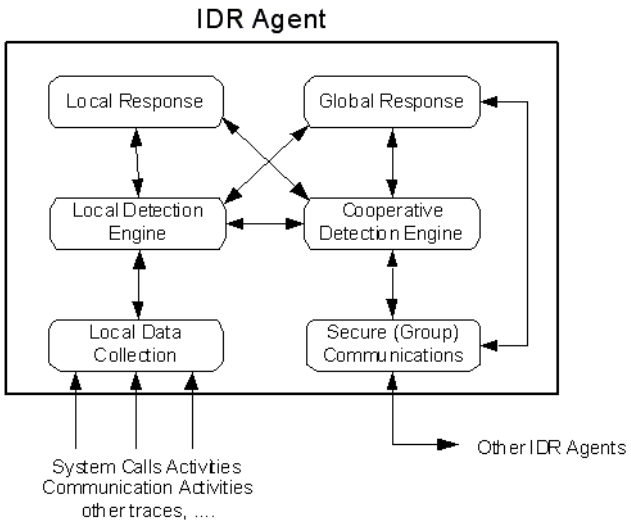

Figure 2: A conceptual model for an IDR agent.

Proposed Work: Figure 2 shows a conceptual model for an Intrusion Detection and Response (IDR) agent similar to the model in [81], but we assume tighter interaction among the modules. Conceptually the IDR agent can be structured into six pieces. The data collection module is responsible for gathering local audit traces, network traffic, and activity logs. The local detection engine will use these data to detect local anomaly and the attacks on the indirect connections relayed by the IDR agent. Detection methods that need broader data sets, or that require collaborations among IDR agents will use the cooperative detection engine. As an example, duplicate messages can be sent over the multiple indirect routes to assist the detection of a comprised proxy server in the SCOLD system. Intrusion response actions are provided by both the local response and global response modules. The local response module triggers actions local to this node, while the global response module realizes the coordinator functions by coordinating actions among IDR agents, such as handling the secure DNS update notification and coordinating the establishment of multiple path secure indirect routing. Finally, a secure group communication module provides a high-confidence communication channel among IDR agents. It coordinates the distribution of group keys for the encryption of packet data over multiple path indirect routes. It also interacts with the cooperative detection engine to re-initialize the communication channels by re-authenticating other IDR agents and reissuing the group key when a compromise node is detected.

We propose to design efficient techniques for tracing the intrusion routes, integrate IDIP with enhanced IDS and adaptive firewall for distributed intrusion detection and handling. Develop specification language for specifying the secure collective defense architecture and the related rules for reconfiguring network and IDS rule updates. For cooperative intrusion push back, we are interested in evaluating QoS related techniques, as described in Section 3.3, can be effective in block DDoS attacks and potential worms from spreading. The multiple indirect routes that is available among the SCOLD participants can be used to exchanged intrusion information for cooperative intrusion detection and handling. Duplicate packets can be sent over different routes as a mechanism for detecting compromised node with in the SCOLD system. We propose to investigate how to utilize the set of proxy servers and the multiple indirect routes in cooperative intrusion detection, isolation, and push back. Efficient layer coding of cyber defense data and secure information sharing will also be investigated.

3.2 Information Fusion

... inputs

3.3 Predictable QoS Differentiation and Regulation

The objective of this work is to extend the traditional on/off IDS model by differentiating the QoS levels based on the confidence about the traffic patterns so as to mitigate the effect of potential cyber attacks on routers and end servers. The idea is similar to Differentiated Services (DiffServ), which was originally proposed and formulated by IETF [15]. Its goal is to define configurable types of packet forwarding in network core routers, which can provide per-hop differentiated services for large aggregates of network traffic. DiffServ has been an active research topic in the arena of packet networks. Many algorithms have been proposed in achieving delay and loss differentiation in the networking core; see [28, 29, 44] for representatives.

In this work, we want to propose innovative resource allocation and scheduling approaches which can quantitatively control the QoS spacings between different traffic classes in terms of response time and slowdown. Figure **??** illustrates the architecture of the work. The admission control will categorize incoming traffic into multiple classes according to their behaviors. The traffic classification can be supported by our network-side efforts. The resource management module will allocate resources for handling the traffic classes differently. Control theory is applied in combination with admission control and resource management for robust QoS differentiation.

Preliminary Results: In [84], we we investigated the problem of processing rate allocation for proportional slowdown differentiation (PSD) on Internet servers. The proportional model states that QoS levels of different traffic classes should be kept proportional to their pre-specified differentiation parameters, independent of the class loads. We first derived a closed form expression of the expected slowdown in an $M/G_P/1$ FCFS queue, which is an $M/G/1$ FCFS queue with a typical heavy-tailed service time distribution (Bounded Pareto distribution). PSD provisioning was realized by deploying a task server for handling each request class in a FCFS manner. We then developed a strategy of processing rate allocation strategy based on the foundations of queueing theory for the task servers in support of PSD provisioning.

Figure 3 shows the percentiles of simulated slowdown ratios of two classes. The upper line is the 95th percentile; the bar is the 50th percentile; and the lower line is the 5th percentile. The simulation results show that under various system load conditions, the proposed rate-allocation strategy can guarantee that the achieved ratios of the average slowdown are close to the corresponding pre-specified differentiation parameter ratios. We find that the strategy can achieve the objective of providing PSD services to different classes in long timescales. The PSD services, however, were provided with large variance. We also found that some requests from class 1 experienced larger slowdowns than those from class 2. This behavior contradicts their pre-specified differentiation ratios (2:1). The results show that the rate-allocation strategy can only provide weak predictability in short timescales. Actually, the strategy acts according to the macro-behavior (class load) of a class rather than its micro-behavior, such as experienced slowdowns of individual requests.

In [82], we proposed a processing rate allocation scheme for providing proportional response time differentiation on Web servers. A challenging implementation issue is how to achieve processing rates for various request classes on individual Web servers. One approach is to divide the total available processes of the server into multiple process pools. Each pool listens to a port and handles the incoming requests. We found that the problem with this *fixed process allocation strategy* is that not all allocated processes are always active due to the workload dynamics. Thus, the ratio of achieved processing rate among classes

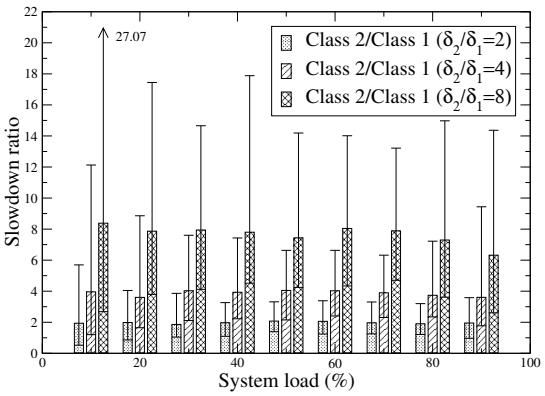

Figure 3: Percentiles of slowdown ratios.

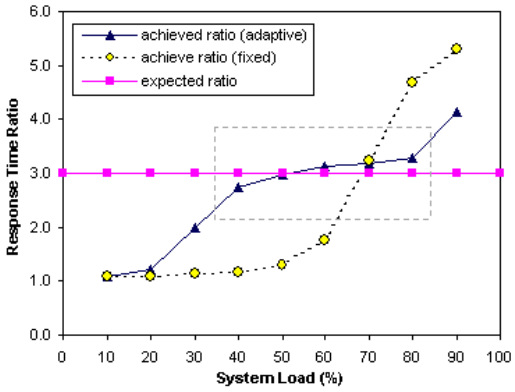

Figure 4: Impact of process allocations.

may not follow the rate allocation scheme. We further proposed an *adaptive process allocation strategy*. Its objective is to adaptively change the number of processes allocated to process pools for handling different classes while ensuring the ratios of allocation specified by the processing rate allocation scheme.

We implemented the two process allocation strategies by modifying child_main() function in http_main.c file on an Apache Web server. The process forking and killing mechanisms were not modified and still handled by Apache. This application-level implementation is hence flexible and portable. We adopted a two-class workload to evaluate the impact of the allocation strategies on the proportional response time differentiation.

Figure 4 shows the achieved response time ratio of two classes due to the two process allocation strategies. It shows that response time differentiation can be achieved with the requests from higher priority classes receiving lower response time than requests from lower priority classes. However, the fixed process allocation strategy cannot achieve proportional response time differentiation because the processing rate of classes cannot be achieved accurately due to the workload dynamics. In contrast, by the use of the adaptive process allocation strategy, when the system load is between 40% to 80%, the difference between the achieved response time ratio and the expected ratio is trivial and hence the proportional response time differentiation is achieved. Figure 4 also shows that when the arrival rate is below 40%, the expected response time ratio is not achieved. This can be explained by the fact that when the workload is light, there is almost no queueing delay observed in all traffic queues. Therefore, DiffServ is not feasible under certain light load conditions, as it was also observed in experiments for delay differentiation in packet networks [29, 44]. When the system load is higher than 80%, we also find out that the expected ratio is not achieved. This can be explained that as the system load is close to its capacity, the impact of the variance of incoming traffic on queueing delay dominates and thus queueing delay in all traffic queues increase significantly. This affects the controllability of the process allocation strategy significantly.

Research Plans: The preliminary results have demonstrated the feasibility of providing predictable QoS differentiation in terms of both slowdown and differentiation by adaptive resource management. The preliminary work advanced our understanding of QoS differentiation techniques to the level where further studies along the line would lead to a breakthrough in the integration of admission control, feedback control, and resource management for providing robust QoS differentiation. In this project, we plan to conduct further studies along the line in the following aspects:

1. Integrating feedback control theory with current queueing theory based resource management so as

to provide QoS differentiation and isolation with smaller variance and better predictability. Feedback control theory was originally developed in physical process control context. In the integrated approach, we plan to employ a queueing theory based resource allocation predicator and a feedback control theory based deviation controller, to provide QoS differentiation at a finer grained level of differentiation predictability and controllability.

2. Integrating admission control with feedback control and adaptive resource management. QoS levels experienced by different traffic classes are based on the differentiated resource allocations and the workload conditions of the classes. Thus, admission control and traffic classification strategies provide another design dimension for providing fine-grained QoS differentiated services. We plan to investigate measurement-based, queueing-theoretical, and control-theoretical admission control strategies, and the integration and interaction of admission control and resource management.

3. Investigating the impact of QoS differentiation for limiting spreading rate of susceptible traffic and regulation traffic under cyber threats.

3.4 Tolerate Intrusion with Proxy-based Multiple-Path Indirect Routes

With the possibility of establishing multiple indirect routes, one can just establish one of the routes, establish multiple indirect routes and then select one of them that meets the QoS or security requirements, or spread packets over the multiple routes dynamically. The existing UDP and TCP protocol only deal with single route. To take advantages of the multiple indirect routes and to facilitate their usage at the application level, we need modify these existing transport protocols or develop new transport protocols. Application programming interface needs to be develop for specifying and controlling the use of these multiple routes. To make smart decision on which route to choose or efficient dynamic scheduling of packets over multiple routes, we need to collect the end to end route quality information. For reliable transport, a TCP session can be establish over each route, but data fragmentation and reassembly needs to be done on top of the multiple TCP sessions to divide and merge the data streams. We can also use a single TCP session but modified the TCP feedback mechanisms to take into consideration of multiple routes. It will be interesting to evaluate which is the right approach.

Preliminary Resules:

Proposed Work: We propose to develop new transport protocol that takes advantages of the available multiple path in the network layers. One design will consist of multiple TCP sessions each utilizes one network path and a thin layer will pro-vide sequencing and reassembly functions. The other design will be to modify existing TCP to handle the spread-ing of packets and flow control on different routes. We will evaluate their performance in the testbed. It is important to investigate how to distribute the data over multiple routes based on desirable requirements such as security, performance, and reliability.

3.5 Evaluation of Cyber-security Systems

In this section we address issues in evaluation of cyber security systems with particular emphasis on predictability issues. As we have moved to the network centric world, the more feared attacks on these same facilities have moved to the cyber world. Unfortunately, it seems the engineering and scientific knowledge developed with years of experience with physical intrusion detection systems, has gone largely unnoticed in the cyber security arena. In particular, while the “sensors” and attacks may be different, the evaluations and methodologies apply in both.

While the recent evaluation work in cyber intrusion detection systems (IDS) such as [65, 64, 4, 46, 45, 47, 49, 3, 21, 67, 77, 5, 8, 7, 9] have made considerable progress, they have yet to consider one of the most basic issues of performance evaluation – *confidence*, at least statistical confidence, in the results. Before we can discuss these works in context, we present a bit more background on detection system evaluation.

To say a system has a detection rate of XX%, which is the state-of-the-art in CIDS, is quite different from saying the probability of detection is YY% with a confidence of 95%. This latter type of statement has been the standard type of requirement (and hence evaluation criterion) for physical intrusion detection system sensors for some time. The former, measurement of a detection rate, is a single sample of performance against a single set (or small set) of conditions. The latter, an estimate of the probability of detection with confidence in that estimate, establishes predictable performance under a range of experimental conditions. As an example of the difference it may help to consider two views of simple coin tosses. If an experiment is run and the DR of "heads" event is 7/10 do we consider the coin "fair"?[1] This is quite different than having a coin where evaluation shows you will get 70% heads, with 95% confidence over any sequence of 10 flips (which is statistically significantly not a fair coin). Measuring DR is easier because it requires only a single "trial", but it also provides very limited "predictability". With the number of attack scenarios, parameters of those scenarios, and variations in background and test, the number of "samples" of any particular setting is often too small for the variations reported to be considered significant.

In the evaluation of a physical intrusion detection system (hereafter called Physical Security Equipment (PSE) to help avoid confusion), the approaches to measure/insure confidence in evaluations have long been studied. The ROC curves, common in the sensor community can, if the measurements are appropriate, allow efficient performance trade-offs, including confidence measures. But the measurements must be of the appropriate type (and numbers) before that can be used. A good example is [6], which discusses the use of statistical process control (SPC), [34], for both initial testing and ongoing assessments of PSE to insure the levels of detection *and confidence* required by DOE regulations for nuclear site security. In [6] both simple (e.g. metal detector) and complex (e.g. hand-geometry based biometric) sensors were discussed.

The SPC approach allows one to approach the confidence estimates in many different ways. [6] highlight some of the differences between testing with binary data (detect/fail), which they call attribute data, and using value data (e.g. event confidence) which is then thresholded to produce a decision (i.e. appropriate ROC type curves). With pure binary data, validating a requirement of 85% probability of detection and 95% confidence level with a single-phase testing attribute testing strategy and allowing no missed alarms required 19 "identical" tests for every "set of conditions" to be considered. Allowing a single miss in the test (so there is a known miss-detection) requires 30 tests per setting to validate a 85% PD with 95% confidence. With value data (ROC curves), *and a stable process*, the number of tests can be significantly reduced down to 5. Note that in both cases the predictability is still limited by the range of test conditions considered and some assumptions of independence, but the repeated tests provide confidence that, if the assumptions are met and the conditions vary according to the experimental parameters, the same results would, 95% of the time, be at the required detection levels.

This model of performance, detection probability with confidence levels, is so standard within the physical intrusion detection community, that government physical security procurement programs routinely write their requirements/specifications in that form without even bothering to provide references to the definitions. To those that work with (proper) ROC curves, their ability to help characterize performance is important. However, as was discussed in [50], the "ROC" curves that have been produced for CIDS systems are not true ROC curves. The values are not actually related to true *probabilities* of detection or false alarms used for standard ROC curves (or even the likelihood ratio test or maximal Bayesian estimator use in other fields).

[1] The probably of 7 or more heads occurring in 10 flips of fair coin is, in fact, 0.17.

Hence they provide little relation to "confidence". Rather the ROCs used in CIDS are detection counts versus false alarms (or worse, false alarm percentages), where the "units" are not well defined. They are closer to the Cumulative Match Score curves used in earlier biometric evaluations[61]. While they can provide some means for comparison, the lack of association with true probability and confidence means no "statically significant" difference can even be discussed.

In the biometric community, many of these issues have not only been addressed in their evaluations, years of research have realized subtle issues that are of particular relevance to CIDS. The underlying assumption in statistical process control or other simple statistical evaluation approaches is that test trials (and implicitly detection failures) are independent and hence that Bernoulli trials and the resulting binomial distributions can be applied to the modeling of the test/failure data to produce confidence intervals. This is, in general, a reasonably good model for "direct" sensors such as magnetic, seismic, vibrational and microwave sensors, where signal and noise are more directly measured and used in the process. However, while it has continued to be used for more complex "detectors", such as biometric", it has presented more difficulties in actually predicting performance.

The SPCA theory is valid, but it is not clear that the assumptions always hold. For face-based biometric systems, we demonstrated the underlying assumptions of Bernoulli were not statistically valid for even moderately large (1000's of trials) sets of biometric data [52]. In particular, we showed that some sets of "individuals" were inherently more difficult to detect/recognize than others and showed a statistically significant "gallery" effect [51]. This fact means that standard Bernoulli-based analysis was not appropriate for producing confidence intervals for this type of data. We also proposed an approach, based on the advanced statistical concepts of balanced repeated replication [71, 68, 42], that, at least over the sample data, allowed us to compute standard errors without assuming independent errors. While at first resisted by others in the biometrics community, within two years other researchers began to realizes that giving up the independence assumptions, while requiring more complex data analysis techniques, allowed more detailed analysis which could focus on what common factors across subsets may make some groups more difficult than others [33]. In the end, we significantly generalized our sampling work to address general sampling design in large scale evaluations, including an approach to detect "unknown co-factors" in biometric evaluation data [54, 55]. The latter question, if co-factors exist, is strongly related to the issues of statistical stationarity and conditional independence of the distribution, and was evaluated on a face-based biometric dataset collection over 6 months that contained over .75TB of data, over 1 Million facial images and 1 Billion biometric signature comparisons. Also discussed in [54], is how any properly normalized ROC, with sufficiently many samples and proper units, can be used to estimate confidence intervals.

Our push to improve the computation of confidence intervals for biometrics has already been adopted by other groups and has been used by NIST for the analysis of FingerPrint systems and the recent Face Recognition Vendor Test 2003 [60, 35, 53].

There have been a few other "evaluation" metrics proposed. Stolfo, [76] proposed a cost based approach, which if the costs were tied to real costs could be very interesting. However quantifying the actual costs is impractical. Even if using a cost model, the uncertainty in the "measurements" are still critical to assessing the "cost risk". In a similar manner, [31] suggests a cost model. That paper mixes costs and ROC analysis, but requires the ROC to be actual probabilities and explicitly states it ignores how to compute an actual (probabilistic valid) ROC. In [77], an evaluation metric, that includes the impact of the response, is considered. Again it is not clear the metric generalizes nor is there any discussion of how statistical variations in the detection would be accommodated in the optimization models.

Another issue commonly acknowledged in the physical intrusion evaluation, which seems to be over-

looked in CIDS literature[2] is the distinction between false alarms and nuisance alarms. While the term comes up in a few publications, none of the evaluation methodologies have a classification for the ground truth that includes nuisance alarms. However, when the data is "scored", it may be very important to separate true false alarms (almost no one would consider a reason for concern) from nuisance alarms (e.g. repeated "stack overflows" by someone who keeps filling out the same "form" data and submitting it, but whose stack data has no "code" in it). The distinction between a nuisance alarm and true alarm can only be knowing the "intent" and in some setting nuisance alarms would be ignored and in others they would want to know. For simulated data, one might argue that nuisance alarms do no occur, but for real traffic it is likely they will.

Having briefly introduced some background from Physical security systems and biometric evaluation, let us now revisit the state of evaluation in CIDS and more general network/cyber threat detection systems.

The majority of the work on CIDS testing and evaluation has focused on the use of simulated attacks, how to effectively generate the attacks and tools for automating the attack process and evaluation, [65, 64, 41, 27, 39, 46, 2]. Use of live data is also common, clearly realistic, but not reproducible. The efficient generation of "test data" is, without question, an important topic. While there will always be arguments about the representativeness of such attack data, [50], it is impractical to have large scale testing that does not use at least moderate amounts of simulated or at least replayed data. We don't expect to add significant new research in "generation" components of these areas. However, to reach meaningful statistical conclusions, we will examine both the proper subdivision of categories of data, and the sampling of that data to achieve statistical confidence (within the range of what can be simulated). From a research perspective this will involve obtaining (or if necessary reimplementing) many of these generation tools, and then developing techniques to more easily reintegrate different mixes of data and then automatically score the results.

We agree with the statements in [8], that new public datasets are needed. While much has been stated about the needs of such a dataset, few have addressed statistical characterization. It is known that different environments have vastly different background behavior, and even the same environment has significant temporal variations. So it is not a single dataset that is needed, but a collection of them. Developing 2-3 different datasets, each with sufficient properties for statistical characterization of performance, is a major goal of our evaluation work.

For datasets that are already existing and labelled, in particular IDEVAL, we will look at two modifications. The first will be to generate "randomizers" that take the original data and perturb the data in various ways to simulate different "samplings" of the original data streams and variations in the timing and order of events. These randomized variations will allow a very restricted form of "confidence" computations from the resulting analysis. While this is less ideal than new generation with proper sampling, many researchers have already used this data, as well as its use in evaluating commercial systems [73]. While it would be good to proceed by extending the best existing protocols/software such as [67], they are, unfortunately, not available for public use.

More "realistic" data has been used in tests conducted by Nophais Labs, [70, 69, 56], with background traffic created by replaying traffic captured from a DePaul University lab. This was useful for high load testing (30-99Mbits/seconds), but no detailed analysis was made of detection rates and it is unclear what ground truth is available. Interestingly, the open-source Snort IDS[66], which we use in our labs, was rated third in their evaluation, outperforming 7 of the 9 commercial products tested. For the more complex datasets, which were a mix of real and/or sanitized traffic and simulated attacks, such as [30, 56], we will explore the variational approach as well, if we can obtain the data.

Given the simulation tools and datasets we will develop hierarchical distributional models, and then test

[2](though hinted at in [50]

the resampled data with the available network-based anomaly detection system to insure they do not find any anomalies. This is necessary because the statistics characteristics may change during re-sampling and we need to find re-samplings that do no change the background "detection" of these systems. In [48], their analysis shows that even the existing (synthetic) IDEVAL data has simulation artifacts that may produce measurable artifacts in anomaly detection systems. In doing so we will also be examining new anomaly detection techniques, based loosely on our prior work in physical intrusion detection and biometrics, that will help insure consistent data generation. (If they don't detect anomalies they will be of little use so we will, of course, test them as detectors as well but that will not be our focus).

While the mixed network-based and host-based evaluations are more complete, and probably more realistic, we will initially focus on the network-based evaluation component. (Testing the host-based often requires obtaining particular versions of host software to exploit, and allow detection of the weakness). Furthermore issues of reinitializing the hosts are more complex, so host-based evaluation will be pursued only after we demonstrate the viability and effectiveness of the enhanced evaluations for network-based evaluations.

There are currently a few automated systems for response to a detected intrusion[63, 75, 13], but most cases the alarm is provided to a system administrator to manually address[19]. The most common is automatically adding firewall rules to block specific sites, but response to DDOS attacks can be considerably more complex involving rate limiting or deflection.

We are unaware of any evaluation that attempts to quantify the impact of mitigation, i.e. evaluating the impact of techniques designed to reduce or eliminate the impact of the attack. The closest would be the work of, [77] which presents a network model and an "algorithm" to evaluate the impact of the response using that model. But that paper presents no significant evaluation or even evaluation methodology. Though it is phrased differently the "cost" models of [76] address some of these issues for standard CIDS and anomaly detection system, but not for complex response models. Furthermore, that work suggested a "cost-based' criterion of evaluation but not an evaluation methodology. How do we effectively evaluate techniques like those proposed in other sections of the proposal that seek to limit the impact with a distributed "response", e.g. one that uses network wide QoS to limit the DDOS, or that uses proxy servers to redirect traffic? We need measures that make sure the responses not only improve local responsiveness (which is all that has been demonstrated to date), but reduce the overall impact of the attack across the full network. (It is not really appropriate for the response to simply make it someone else's problem). These are research questions that will be addressed, in part by instrumented toolsets and in part, we expect, by designing special experiments to measure the requested QoS or routing changes and then using quantitative network models to estimate the impact of such changes.

As more and more groups develop adaptive systems, adaptive in both their "detection" technology and their response, testing/evaluation with canned data becomes impractical. As "live" evaluations becomes required, it is important to design not just for benchmark datasets, but for on-line evaluations. We note that comparative evaluations with differing backgrounds require significantly more data and analysis to show any type of statistical significance. However, this need is mollified by the knowledge that testing with more realistic background variations, and the requisite larger amounts of testing, will also provide for predictive behavior over a wider range of situations. But still we need to worry about a balanced comparison. To address this, the tool is being designed to provide multiple simultaneous "feeds" for live evaluation, with the instrumentation needed for evaluation of the results computed for each system in parallel. Statistically similar attack profiles will be run simultaneously. Responses to probes from the active systems will be "answered" independently for each of the simulated attacks, but live traffic from the different systems will be mixed for responses to live nodes.

Because the amount of testing needed to reach statistical confidence will be larger than current benchmarks have used, it is anticipated that total automation will be required. While host-based techniques are not expected to be our focus, we still need to have the ability to "restore" to a clean host and then continue the evaluation. The automation is expected to use VMware to support, at least for Windows^{TM} and Linux target OSs, an automated "boot/run/infect" sequence. We will explore, in year 2, how easily this will allow us to include host-based attacks in the evaluation toolset.

In summary, to support the dataset generation and analysis we will be developing the IMPACT toolset, a distributed cyber-evaluation toolset which will facilitate the parallel dumping, sanitizing and scoring of network traffic, supporting generation of new datasets as well as on-line evaluations. It will work as a coordinated distributed systems to allow processing/collecting data at closer to backbone speeds, but with added time-stamping and indexing to support enhanced playback capabilities necessary for randomization to support "statistical" confidence testing. The tool chain will contain components to allow it to programmatically "remap" users, IP addresses and even some content. (The exact details of the remapping is unclear, but we will do what it takes to satisfy our university IT staff we can maintain privacy while trying to maintain the fidelity of the original data).

The IMPACT toolset draws on our experience in biometric system evaluation and has 4 primary goals:

1. develop collections and playback that support the full range of resampling we believe is necessary for effective statistical evaluation.
2. support multi-host (in parallel) simultaneous online evaluation of network-based CIDS.
3. automated "scoring" based on the live data and/or playbacks to include measuring statistical confidence.
4. support evaluation of intrusion mitigation technologies

4 Broader Impact of the Project

In additional to the technical contributions of the research, this proposal will have 2 significant broader impacts.

The first is a mixture of education and societal. As part of our program's outreach to the community, we have arranged to develop a local cable-television show focused on cyber security issues within our community. The show, minimally 30min per month and possibly more depending on costs/sponsorship and viewer feedback, will seek to educate, and actively involve the local IT workforce in cyber security issues. The Colorado Springs area has almost 200,000 white-collar and military employees. The surrounding areas, some of which are served by that same cable company and would receive the broadcasts, nearly doubles that level. The area is home to multiple Military groups including US Northern Command (in charge of US Military Homeland Defense), Cheyenne Mounting Complex (US Strategic Command), Fort Carson, Peterson AFB, Shriver AFB, and the Air Force Academy. There are significant installations of major corporations including Intel, Amtel, LockHeed-Martin, Raytheon, ITT, HP/Compaq, Boeing, MCIworldcom, Quantum, Oracle, Federal Express Data Systems, Adelphia Cable, Allied Signal, Qwest Communications, Comp.Sci.Corp., TRW, DRS, Gateway 2000, as well as three major hospitals and a significant school system. By making it local and related to things they know, we expect this TV show to help to keep this high-tech workforce more up to date and significantly improve our local impact. This effort will be pursued in conjunction with our Networking Information and Space Security Center in Colorado Springs, which

already has significant ties with Northern Command and the local military which, given their missions, are very interested in cyber security issues. In conjunction with NISSC, the UCCS CS Department is supporting a certificate program in Information Assurance, some of which might be used to add more detailed technical content to the broadcasts (if there is sufficient demand for the increased frequency and increased depth).

The second impact will be on the university itself, where it will be engaging both graduate and undergraduate students. In its 2004 college rankings edition, "America's Best Colleges," US News's editors ranked CU-Colorado Springs 5th among public master's universities in the West. While the school's history is one of undergraduate and MS level education with pockets of research, it is on the road to becoming a regional research university. It has had doctoral programs in Engineering for over a decade, but only small amounts of funded research – limiting the growth of the doctoral program. Dr. Chow has been doing productive research at UCCS for years, but with little funding. Dr. Zhou is an assistant professor who joined UCCS in August 2003. Dr. Boult also joined in August 2003, and chose to move to UCCS (from Lehigh) in part because of the chance to provide research opportunities where few existed, as well as to reach a different group of students. The proposal will help fund new doctoral students and aid in the transformation of UCCS to a regional research university. UCCS has a non-traditional undergraduate population with a mostly commuting student body, a median undergraduate age approaching 30. A majority of the UCCS students are self-supporting, and opportunities to work on campus will improve their chances of success and potential to have research or advanced development careers. This funding will provide unique opportunities to these students.

5 Research Schedule and Deliverables

The goal of this project is to develop techniques for improving measurable performance of network system under cyber attacks. Specifically, we are going to:

1. Develop intrusion toleration techniques based on autonomous establishment of secure multiple indirect routes for legitimate connections to mitigate the impact of inevitable cyber attacks. DNS system will be improved and enhanced with the support of multiple indirect routes.

2. Develop Transport protocols that utilize proxy server based multiple routes will developed together with API for specifying and controlling of multiple indirect routes.

3. Develop effective cooperative intrusion detection and response systems and explore the use of multiple indirect routes for distributed intrusion detection, network reconfiguration, and compromised node detection.

4. Integrate admission control with adaptive resource management mechanisms for QoS differentiation and isolation for limiting the spreading rate of susceptible traffic and mitigating effect of 3DoS attacks. Admission control mechanisms based on traffic measurement and pattern recognition admit and classify incoming traffic into multiple classes with different confidence levels. QoS-adaptive resource management mechanisms with feedback control provide robust QoS differentiation and isolation to the multiple classes.

5. Develop the IMPACT tool-set, which will be a a distributed cyber-evaluation tool-set which will facilitate the parallel dumping, sanitizing and scoring of network traffic and thus support generation of new datasets as well as on-line evaluations. The tool-set will log with added time-stamping and indexing

to support enhanced playback capabilities necessary for the resampling techniques we propose to support "statistical" confidence testing. The tool chain will support mapping tables to "remap" users, IP addresses and even some content, thus supporting enhanced resampling as well as addressing privacy issues. The result will be both a tool-set and new significant public datasets for evaluations.

The research team will include the two PIs, three graduate research assistants, and three undergraduate research assistants. Dr. Chow has extensive experience in network and protocol design, network restoration, content switching, and network security. Dr. Zhou's expertise is in QoS-adaptive scheduling and resource management on distributed systems. We are teaming up to investigate cost-effective solutions for establishing trustworthy network system performance even under cyber attacks.

One GRA will be working to improve secure indirect routing and develop the enhanced transport protocols and the related API that utilize multiple routes, with one semester developing the proxy server selection algorithms, two semesters enhancing secure indirect routing and secure DNS update, and three semesters on designing and implementing multi-route transport protocols and the relate API. One GRA will develop the cooperative intrusion detection and response system, with two semesters on developing the IDIP protocol with enhancement for the cooperative push back, four semesters on the cooperative detection and isolation. For the development of the IMPACT tool-set for evaluation, there will be one GRA working on it and 1-2 undergraduates working part-time when we do the evaluations. In keeping with past experience we also expect 2-3 MS level students working on MS thesis level research on the evaluation methodology and particular tools. These unpaid MS students are still expected to produce conference/workshop papers and their travel expenses may come from the grant. The student(s) working on IMPACT are expected to interact closely with, and may be contributing to the other aspects of the project, especially the distributed detection components. The fourth GRA will be working on the modeling and analysis of the integration of admission control and feedback control with adaptive resource management for providing predictable QoS levels on network routers and servers. One undergraduate assistant will be working on the implementation and evaluation of the proposed algorithms and mechanisms. The proposed research is planned to be carried out on an available experimental testbeds at the Networking and Systems Laboratory, with which the PIs are affiliated.

The deliverables include the publications of research results, the software packages developed for enhanced secure DNS update, multiple indirecting routing, and cooperative IDR system, a library of the integration of admission control, feedback control, and resource management algorithms, and technical reports after each milestone. The reports will be published in ACM/IEEE sponsored leading technical conferences and journals. Any software package resulted from this project will be released through the project homepage for the public use free of charge.

6 Conclusion

This project, will make important contributions to improving the measurable performance against cyber-attacks, addressing all four key elements, detection, local mitigation techniques, network wide QOS mitigation, and evaluation. It is expected to make contributions on theoretical model, practical experimental systems, toolsets for general use, and detailed evaluation datasets. The educational impact of this proposal at UCCS, its interaction with the Network Information and Space Security Enter, and the in the large white-collar workforce in and around Colorado Springs will be significant.

References

[1] T. F. Abdelzaher, K. G. Shin, and N. Bhatti. Performance guarantees for Web server end-systems: a control-theoretical approach. *IEEE Trans. on Parallel and Distributed Systems*, 13(1):80–96, 2002.

[2] S.J. Aguirre and W.H.Hill. Intrusion detection fly-off: Implications for the united states navy. Technical report, MITRE, 1997.

[3] D. Alessandri. Using rule-based activity descriptions to evaluate intrusion detection systems. In H. Debar, L. Me, and S. F. Wu, editors, *Int. Workshop on Recent Advances in Intrusion Detection*, volume 1907 of *Lectures in CS*, pages 183–196. Springer Verlag, 2000.

[4] E. Amoroso and R. Kwapniewski. A selection criteria for intrusion detection systems. In *Proc. 14th AnnualComputer Security Applications Conference*, pages 280 – 288. IEEE, Dec 1998.

[5] G.A. Fink and B.L. Chappell and T.G. Turner and K.F. O'Donoghue. A metrics-based approach to intrusion detection system evaluation for distributed real-time systems. In *Proc. Int. Parallel and Distributed Processing Symposium*, pages 93–100. IEEE, 2002.

[6] D.W. Murray and D.D. Spencer. Statistical process control testing of electronic security equipment. In *IEEE Int. Carnahan Conf on Security Technology*, pages 53–59. IEEE, Oct 1994.

[7] W.H. Allen and G.A. Marin. On the self-similarity of synthetic traffic for the evaluation of intrusion detection systems. In *Proc. Symp. on Applications and the Internet*, pages 242–248. IEEE, Jan 2003.

[8] N. Athanasiades and R. Abler and J. Levine and H. Owen and G. Riley. Intrusion detection testing and benchmarking methodologies. In *First IEEE Int. Workshop on Informatoin Assurance (IWIAS)*, pages 63–72. IEEE, March 2003.

[9] K.M.C. Tan and R.A. Maxion. Determining the operational limits of an anomaly-based intrusion detector. *IEEE Journal on Selected Areas in Communications*, 21(1):96–110, Jan 2003.

[10] E. et al Ayanoglu. Diversity coding for transparent self-healing and fault-tolerant communication networks. *IEEE Trans. on Communications*, 41(11), 1993.

[11] A. Bakre and B. R. Badrinath. I-tcp: Indirect TCP for mobile hosts. In *Proc. 15th Int. Conf. Distributed Computing System (ICDCS)*, 1995.

[12] H. Balakrishnan, S. S. Seshan, , and R. Katz. Improving reliable transport and handoff performance in cellular wireless networks. *Wireless Network*, 1(4), 1995.

[13] I. Balepin, S. Maltsev, J. Rowe, and K Levitt. Using specification-based intrusion detection for automated responce. In *Proc. International Symp on Rapid Advances in Intrusion Detection (RAID)*, Pittsburg, PA, USA, September 2003.

[14] N. Bhatti and R. Friedrich. Web server support for tiered services. *IEEE Network*, 13(5):64–71, 1999.

[15] S. Blake, D. Black, M. Carlson, E. Davies, Wang Z., and W. Weiss. An architecture for differentiated services. *IETF RFC 2475*, 1998.

[16] L. Briesemerister, P. Lincoln, and P. Porras. Epidemic profiles and defense of scale-free networks. In *Proc. of ACM WORM*, 2003.

[17] K. Brown and S. Singh. M-TCP: TCP for mobile cellular networks. *ACM Comput. Commun.*, 27(5):19–43, 1997.

[18] D. W. Carman, P. S. Kruus, and B. J. Matt. Constraints and approaches for distributed sensor network security. Technical report, NAI Labs Technical Report 00-010, 2000.

[19] C. A. Carver, J. M. D. Hill, and U. W. Pooch. Limiting uncertainty in intrusion response. In *Proc. of IEEE Workshop on Information Assurance and Security*, US Military Academy, West Point, June 2001.

[20] A. Cearns and C. E. Chow. A2d2: Design of an autonomous anti-ddos (a2d2) network. In *Proc. of IASTED Conf. on Applied Informatic*, 2003.

[21] T. Champion and M. Denz. A benchmark evaluation of network intrusion detection systems. In *Proc. of IEEE Conf. on Aerospace Systems*, 2001.

[22] S. Chandra, C. S. Ellis, and A. Vahdat. Application-level differentiated multimedia Web services using quality aware transcoding. *IEEE J. on Selected Areas in Communications*, 18(12):2544–2265, 2000.

[23] J. Chen. New approaches to routing for large scale data networks. Technical report, Ph.D. Dissertation, Rice University, 1999.

[24] S. Chen, S. Cheung, R. Crawford, and M. Dilger. GrIDS-a graph based intrusion detection system for large networks. In *In Proc. of the 19th National Information Systems Security Conference*, 1996.

[25] X. Chen and P. Mohapatra. Performance evaluation of service differentiating Internet servers. *IEEE Trans. on Computers*, 51(11):1,368–1,375, 2002.

[26] C. E. Chow, P. J. Fong, and G. Godavari. An exercise in constructing secure mobile Ad Hoc networks. In *Proc. of Int'l Conf. on Advanced Information Networking and Applications*, 2004.

[27] K. Das. The development of stealthy attacks to evaluate intrusion detection systems. Master's thesis, MIT EECS, June 2000.

[28] C. Dovrolis, D. Stiliadis, and P. Ramanathan. Proportional differentiated services: Delay differentiation and packet scheduling. In *Proc. ACM SIGCOMM*, 1999.

[29] C. Dovrolis, D. Stiliadis, and P. Ramanathan. Proportional differentiated services: Delay differentiation and packet scheduling. *IEEE/ACM Trans. on Networking*, 10(1):12–26, 2002.

[30] National Laboratory for Applied Network Research. Nlar network traffic packet header traces, 2002. http://pma.nlanr.net/Traces/.

[31] J.E. Gaffney and J.W. Ulvila. Evaluation of intrusion detectors: A decision theory approach. In *IEEE Symp. on Security and Privacy*, Oakland, CA, May 2001. IEEE.

[32] M. Garetto, W. Gong, and D. Towsley. Modeling malware spreading dynamics. In *Proc. of IEEE INFOCOM*, 2003.

[33] G. Givens, J.R. Beveridge, B.A. Draper, and D. Bolme. A statistical assessment of subject factors in the pca recognition of human faces. In *"IEEE Workshop on Statistical Analysis in Computer Vision*. IEEE, June 2003.

[34] E.L. Grant and R.S Leavenworth. *Statistical Quality Control*. McGraw-Hill, 1972.

[35] P.J. Grother, R.J. Micheals, and P. J. Phillips. Face recognition vendor test 2002 performance metrics. In *Proceedings 4th International Conference on Audio Visual Based Person Authentication*, June 2003.

[36] Liu. J. and S. Singh. Atcp: Tcp for mobile ad hoc networks. *IEEE J. on Selected Areas on Communications*, 19:1300–1315, 2001.

[37] Klaus Julisch. Clustering intrusion detection alarms to support root cause analysis. *ACM Trans. on Information and System Security*, 6(4):443–471, 2003.

[38] A. Karlof and D. Wagner. Secure routing in sensor networks: Attacks and countermeasures. In *Proc. of 1st IEEE Int'l Workshop on Sensor Network Protocols and Applications*, 2003.

[39] K. Kendall. A database of computer attacks for the evaluation of intrusion detection systems. Master's thesis, MIT EECS, 1999.

[40] B. Ko, K. Lee, K. Amiri, and S. Calo. Scalable service differentiation in a shared storage cache. In *Proc. 23rd IEEE Int'l Conf. on Distributed Computing Systems (ICDCS)*, 2003.

[41] J. Korba. Windows nt attacks for the evaluation of intrusion detection systems. Master's thesis, MIT EECS, June 2000.

[42] D. Krewski and J. N. K. Rao. Inference from statified samples: Properties of the linearization, jackknife and balanced repeated replication methods. *The Annals of Statistics*, 9(5):1010–1019, 1981.

[43] S. C. M. Lee, J. C. S. Lui, and D. K. Y. Yau. Admission control and dynamic adaptation for a proportional-delay DiffServ-enabled Web server. In *Proc. ACM SIGMETRICS*, 2002.

[44] M. K. H. Leung, J. C. S. Lui, and D. K. Y. Yau. Adaptive proportional delay differentiated services: Characterization and performance evaluation. *IEEE/ACM Trans. on Networking*, 9(6):908–817, 2001.

[45] R. Lippmann, J. Haines, D. Fried, J. Korba, and K. Das. The 1999 darpa off-line intrusion detection evaluation. *Computer Networks*, 34:579–595, 2000.

[46] R. P. Lippmann, D. J. Fried, I. Graf, J. W. Haines, K. P. Kendall, D. McClung, D. Weber, S. E. Webster, D. Wyschogrod, R. K. Cunningham, and M. A. Zissman. Evaluating intrusion detection systems: The 1998 darpa off-line intrusion detection evaluation. In *Proceedings of the 2000 DARPA Information Survivability Conference and Exposition*. DARPA, 2000.

[47] R. P. Lippmann, J. W. Haines, D. J. Fried, J. Korba, and K. Das. The 1999 darpa off-line intrusion detection evaluation. Technical report, MIT Lincoln Lab, 2000.

[48] M.V. Mahoney and P.K. Chan. An analysis of the 1999 darpa/lincond laboratory evaluation data for network anomaly detection. In *Proc. Recent Advances in Intrusion Detection*, volume 2820 of *Lectures in CS*, pages 220–237. Springer Verlag, November 2003.

[49] R.A. Maxion and K.M.C. Tan. Benchmarking anomaly-based detection systems. In *IEEE Proc Int. Conf on Dependable Systems and Networks*, pages 623–630, 2000.

[50] J. McHugh. Testing intrusion detection systems: A critique of the 1998 and 1999 darpa off-line intrusion detection system evaluation as performed by lincoln laboratory. *ACM Transactions on Information and System Security*, 3(4), November 2000.

[51] R. J. Micheals and T. E. Boult. Efficient evaluation of classification and recognition systems. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2001)*, Hawaii, December 11–13 2001.

[52] R. J. Micheals and T. E. Boult. A stratified methodology for classifier and recognizer evaluation. In *IEEE Workshop on Emperical Evaluation Methods in Computer Vision*, Kauai, Hawaii, Dec 2001. IEEE.

[53] R. J. Micheals, P. Grother, and P.J. Phillips. The nist human id evaluation framework. In *Proc. of the 4th Int. Conference on Audio and Video-based Biometric Person Authentication*, June 2003. Availible from http://www.frvt.org/DLs/AVBPA-2003.pdf.

[54] R.J. Micheals. *Biometric systems evaluation*. PhD thesis, Lehigh University, 2003.

[55] R.J. Micheals and T.E. Boult. Is the urn well-mixed? Technical report, National Institute of Standards and Technology, Febuary 2004.

[56] P. Mueller and G. Shipley. Dragon claws its way to the top. *Network Computing*, August 2001. http://www.networkcomputing.com/1217/1217f2.html.

[57] Boeing Phantom Works. Network Associates Labs. Intrusion detection and isolation protocol, IDIP. Technical report, 2002.

[58] P. Ning, S. Jajodia, and S. Wang. Abstraction-based intrusion detection in distributed environments. *ACM Trans. on Information and System Security (TISSEC)*, 4:407–452, 2001.

[59] A. Perrig, R. Szewczyk, J.D. Tygar, Victorwen, and D. E. Culler. Spins: Security protocols for sensor networks. *Wireless Networks*, 8:521–534, 2002.

[60] P.J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and M. Bone. Face recognition vendor test 2002. evaluation report. Technical Report IR 6965, National Institute of Standards and Technology, March 2003. www.itl.nist.gov/iad/894.03/face/face.html.

[61] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, October 2000.

[62] P. A. Porras and P. G. Neumann. Emerald: event monitoring enabling responses to anomalous live disturbances. In *In 1997 National Information Systems Security Conference*, 1997.

[63] P. A. Porras and P. G. Neumann. Emerald: Event monitoring enabling responses to anomalous live disturbances. In *Proceedings of the 20th NIS Security Conference*, October 1997.

[64] N. Puketza, M. Chung, R. A. Olsson, and B. Mukherjee. A software platform for testing intrusion detection systems. *IEEE Software*, pages 43–51, September/October 1997.

[65] N. J. Puketza, K. Zhang, M. Chung, B. Mukherjee, and R. A. Olsson. A methodology for testing intrusion detection systems. *IEEE Transactions on Software Engineering*, 22(10), 1996.

[66] M. Roesch. Snort - lightweight intrusion detection for networks. In *USENIX 13th Systems Administration Conference - LISA '99*, Seattle, Washington, 1999. usenix. see also www.snort.org.

[67] L.M. Rossey, R.K. Cunningham, D.J. Fried, J.C. Rabek, R.P. Lippmann, J.W. Haines, and M.A. Zissman. Lariat: Lincoln adaptable real-time information assurance testbed. In *IEEE Proc. Aerospace Conference*, volume 6, pages 2671–2682, March 2002.

[68] J. Shao and C. F. J. Wu. Asymptotic properties of the balanced repeated replication method for sample quantiles. *Annals of Statistics*, 20(3):1571–1593, September 1992.

[69] G. Shipley. Intrusion detection, take two. *Network Computing*, November 1999. http://www.networkcomputing.com/1023/1023f1.html.

[70] G. Shipley. Iss realsecure pushes past newer ids players. *Network Computing*, May 1999. http://www.networkcomputing.com/1010/1010r1.html.

[71] R. R. Sitter. Balanced repeated replications based on orthogonal multi-arrays. *Biometrika*, 80(1):211–221, March 1993.

[72] S. Snapp, J. Brentano, and G. Dias. Dids (distributed intrusion detection system) motiva-tion, architecture, and an early prototype. In *In Proceedings of the 14th National Computer Security Conference*, 1991.

[73] D. Song, G. Shaffer, and M. Undy. Nidsbench - a network intrusion detection test suite. In *Recent Advances in Intrusion Detection, Second International Workshop*, West Lafayette, 1999. http://www.raid-symposium.org/raid99/PAPERS/Song.pdf.

[74] E. H. Spafford and D. Zamboni. Intrusion detection using autonomous agents. *Computer Networks*, 34(4):547–570, 2000.

[75] D. Sterne, K. Djahandari, B. Wilson, B. Babson, D. Schnackenberg, H. Holliday, and T. Reid. Autonomic response to distributed denial of service attacks. In *Proc. International Symp on Rapid Advances in Intrusion Detection (RAID)*, Davis, CA, USA, October 2001.

[76] S. Stolfo, W. Fan, W. Lee, A. Prodromidis, and P. Chan. Cost-based modeling for fraud and intrusion detection: Results from the jam project. In *n Proceedings of the 2000 DARPA Information Survivability Conference and Exposition (DISCEX '00)*, 2000.

[77] T. Toth and C. Kruegel. Evaluating the impact of automated intrusion response mechanisms. In *18th Computer Security Applications Conference*, pages 301–310. IEEE, December 2002.

[78] A. Tsirigos and Z. J. Haas. Multiple path routing in the presence of frequent topologicalchanges. *IEEE Communication Magzine*, pages 132–139, 2001.

[79] R. Yavatkar and N. Bhagawat. Improving end-to-end performance of TCP over mobile internetworks. In *Proc. IEEE Workshop on Mobile Computing Systems and Applications*, 1994.

[80] M.G. Zapata. Secure Ad Hoc on-demand distance vector (SAODV) routing. Technical report, http://www.ietf.org/internet-drafts/draft-guerrero-manet-saodv-00.txt, Internet Draft, 2001.

[81] Y. Zhang, W. Lee, and Y. Huang. Intrusion detection techniques for mo-bile wireless networks. *Wireless Network*, 9:545–556, 2003.

[82] X. Zhou, Y. Cai, G. K. Godavari, and C. E. Chow. An adaptive process allocation strategy for proportional responsiveness differentiation on Web servers. In *Proc. IEEE 2nd Int'l Conf. on Web Services (ICWS)*, July 2004.

[83] X. Zhou, J. Wei, and C.-Z. Xu. Modeling and analysis of 2D service differentiation on e-Commerce servers. In *Proc. of IEEE 24th Int'l Conf. on Distributed Computing Systems (ICDCS)*, pages 740–747, March 2004.

[84] X. Zhou, J. Wei, and C.-Z. Xu. Processing rate allocation for proportional slowdown differentiation on Internet servers. In *Proc. IEEE 18th Int'l Parallel and Distributed Processing Symp. (IPDPS)*, pages 88–97, April 2004.

[85] X. Zhou and C.-Z. Xu. Harmonic proportional bandwidth allocation and scheduling for service differentiation on streaming servers. *IEEE Trans. on Parallel and Distributed Systems*, 15(9):835–848, 2004.

[86] H. Zhu, H. Tang, and T. Yang. Demand-driven service differentiation for cluster-based network servers. In *Proc. IEEE INFOCOM*, pages 679–688, 2001.