PROJECT DESCRIPTION

1 Introduction

...integration of the motivations & objectives.

1.1 Motivations

(Architecture by Chow)

Figure 1 shows the Secure Collective Defense (SCOLD) architecture and illustrates how secure indirect route can be achieved in such system. We assume each SCOLD site contributes a SCOLD Intrusion Detection and Response (IDR) Agent. Each SCOLD IDR agent consists of a SCOLD Proxy server, several gateways, and a SCOLD coordinator. It is assumed that multiple alternate gateways that can be pre-deployed or established dynamically during the cyber attacks. When the cyber attacks occur, the local intrusion detection system notifies the SCOLD coordinator to initiate secure DNS updates and coordinate the secure indirect route set up. We also assume that a collection of geographical SCOLD proxy servers is available in Internet either contributed by the participants of a SCOLD consortium, branches of an organization, or a feebased network service providers. Each proxy server will be assigned to update the DNS server of a client network through a secure DNS update protocol, or to update directly the routing table of specific clients with the pre-arranged agreement. The secure DNS update includes new DNS entries with multiple indirect routing information, i.e., besides the normal domain name, IP address, it includes a set of designated proxy servers which a client can use to establish indirect route(s) to the target site. Client DNS server can query the target DNS via the indirect route when the main route are congested or under cyber attacks. On receiving the new DNS query information, the client can establish a IP tunnel through the selected proxy server to the target site. For example, here SCOLD proxy B and SCOLD proxy C are chosen and two indirect routes are setup via two different alternate gateways, Gateways 2 and 3. The set of SCOLD IDR agents form a wide area demilitarized zone (DMZ) to protect the participants of the SCOLD system. Further attacks to the indirect routes can be blocked by a SCOLD proxy server with an integrated intrusion detection and handling system.

Figure 2 shows a conceptual model for an Intrusion Detection and Response (IDR) agent similar to the model in [105], but we assume tighter interaction among the modules. We assume each SCOLD site or the SCOLD proxy server will be realized with such IDR agent. 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 SCOLD 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

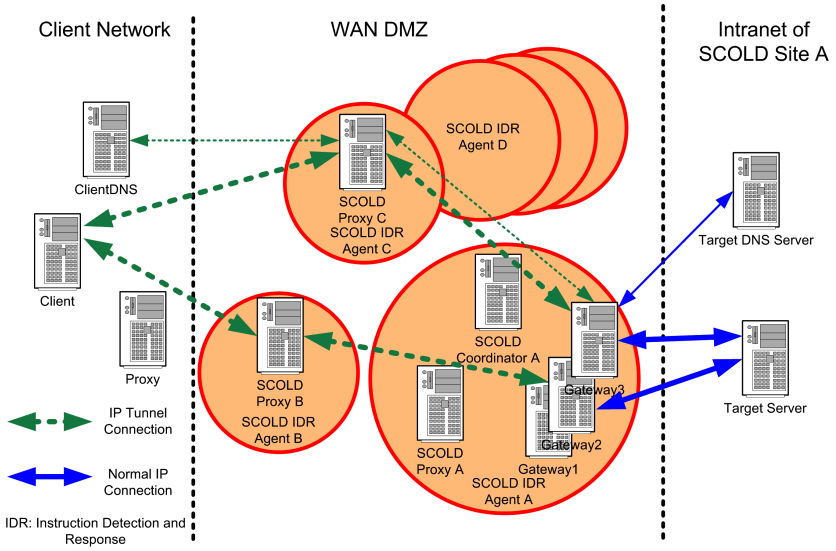

Figure 1: The architecture of the Secure Collective Defense System.

the communication channels by re-authenticating other IDR agents and reissuing the group key when a compromise node is detected.

(By Joe) As more and more mission-critical services are provided on the Internet, the threats to the services are increasing also. The past years have witnessed the tides of Distributed Denial of Service (DDoS) attacks on high-profile Web sites. We are also experiencing disturbingly frequent outbreak of malicious worms and viruses (so called *malware*) [19, 41]. As we more and more rely on the intrusion detection systems (IDS), the IDS systems also give us more and more false alarms due to the new attack patterns. Therefore, the traditional on/off IDS model is not sufficient and effective. We want to extend the functionalities of IDS model by providing more options. For example, if the confidence of an intrusion is below a certain threshold, the system may use QoS degrading based on the confidence as a means to limit the propagation rate of the potential malicious and susceptible traffic.

Current DDoS attacks aims to completely disable the victim system's service to its clients by consuming its available resources. We call it disruptive attacks, some typical defense mechanisms are based on IP traceback techniques to locate the attackers and hence shut down them through administrative means. Compared to the 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 quality of service (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, current defense mechanisms may not be effective or sufficient under 3DoS threats. QoS differentiation is a feasible way to control the resource consumption under such susceptible 3DoS attacks.

We want to integrate admission control strategies with QoS-adaptive resource management mechanisms on network routers and servers to extend the essential on/off model in current IDS systems. The basic idea is to control the access and consumption of clients to network system resources based on their behaviors and the confidence level the system has for making the QoS regulation decisions. The mechanisms will guarantee premium QoS to legitimate and well-behaving clients. Aggressive clients, who may be potentially malicious or legitimate, will be given fair QoS when the server workload is low and their perceived QoS will be

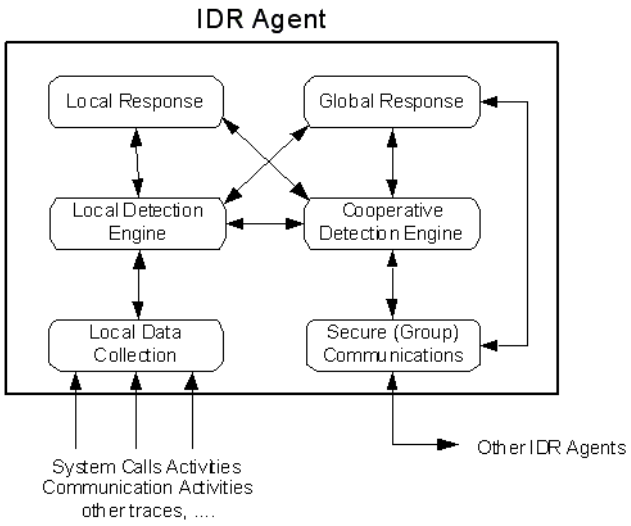

Figure 2: A conceptual model for an IDR agent.

degraded as their traffic increases. Thus, the strategy is to regulate traffic by QoS differentiation and isolation with admission control and resource management at the server side and it does not rely on IP traceback techniques. Our research is driven by opportunities coming from service differentiation techniques, which aim to provide predictable and controllable QoS levels to different clients based on clients' access patterns and servers' resource capacities. The uniqueness of our work lies in the integration of admission control with adaptive resource management for robust QoS differentiation so as to extend the functionalities of the traditional on/off IDS model.

1.2 Objectives

The goal of this project is ... Specifically, we are going to:

1.

2.

3.

4. Integrate admission control with adaptive resource management mechanisms for QoS differentiation and isolation for 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 QoS expectations. QoS-adaptive resource management mechanisms with feedback control provide robust QoS differentiation and isolation to the multiple classes under 3DoS attacks. The integration aims to mitigate the effect of 3DoS attacks by regulating the movement of traffic and hence processing differently.

5. (by Boult...evaluation of cyber-security systems...)

2 Proposed Work

2.1 Tolerate Intrusion with Proxy-based Multiple Path Secure Indirect Routing

Intrusion tolerance paradigm explores techniques to allow legitimate users to communicate even under DDoS attacks. The goals are to maintain certain level QoS for the legitimate connections while keeping the processing overhead as low as possible under cyber attacks. It supplements the intrusion push back mechanism by providing assurance with additional measurable performance. Various request redirection and adaptive QoS-based queueing techniques were proposed for intrusion tolerance. However many of them are aimed at improving the performance of connections via direct Internet routes. Under severe DDoS attacks, such techniques suffer significant performance degradation. In this proposed research, we will focus on techniques for establishing multiple securing indirect routes between legitimate users over a set of proxy servers and alternate gateways. It is reason-able to assume that alternate gateways exist or can be dynamically established for supporting indirect routes under cyber attacks. It is also feasible to find a set of proxy servers at geographical diverse locations for relaying packets over the indirect routes. Those proxy servers can be contributed by the branches of a company or a consortium of organizations. They can also be provided with fees by a commercial network service company. Since it is difficult to tell which users are comprised, the proposed techniques need to deal further attacks on those proxy servers and alternate gateways.

We propose to design a new secure DNS protocol that supports the secure update and query of new DNS entries with the multiple indirect route information and to develop secure indirect routing protocol that utilizes the DNS query result with multiple indirect route information. The availability of these multiple secure indirect routes offers the following benefits:

1. **Security**. When a route gets attacked, the users can switch to a different route dynamically. Urgent or critical packets can be sent over multiple routes to increase their chance of delivery. Encrypted content can be spread over multiple paths and make it harder to decrypt. The information of further DDoS attacks on the proxy servers can be used to identify and isolate the source of the attack.

2. **Reliability**. Users can choose a reliable route among the direct route and a set of indirect routes. Packet content can be spread over multiple routes with redundant information or forward error correction code to reduce the packet loss or corruption.

3. **Performance**. The multiple indirect routes provide additional bandwidth that is not available with exist-ing Internet direct routing. It can be used for dynamic bandwidth provisioning and satisfy the unexpected or sudden bandwidth needs.

Figure 3 shows the victim under DDoS attack, without the alternate routes. The bandwidth of legitimate clients is greatly reduced due to large DDoS attack traffic volume and the nature the first in first out routing mechanism. Figure 4 shows the situation with the implementation of secure indirect routes. Since some clients may be compromised as attack agent, we will not share the alternate gateway information directly with clients. A set of proxy servers is recruited to form a wide area DMZ. The clients will get updated alternate DNS entry in-formation with the designated proxy server and their packets will be redirected by the proxy server through the designated alternate gateway to the final destination. Only the designated proxy server will be informed about the corresponding alternate gateway and what are subnets to be accepted. The proxy server will be enhanced with integrated IDS and adaptive firewall to block further attack traffic.

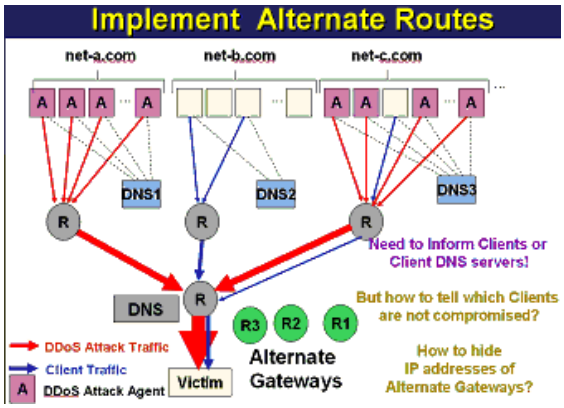

Figure 3: A system without alternate routes.

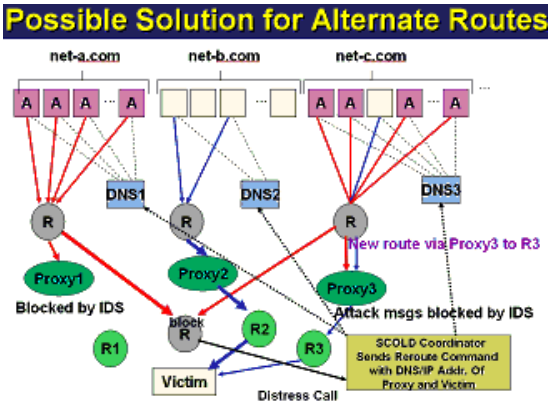

Figure 4: A system with secure indirect routing.

Previous Work/Preliminary Work: Berkeley Internet Name Domain (BIND) server software, originally written by Kevin Dunlap for the Berkeley 4.3 BSD UNIX operating system [2]. BIND is currently the de facto standard and is used by most UNIX operating systems as well as Windows NT [2]. It is maintained by the Internet Software Consortium (ISC) and can be downloaded from ISC's web site at www.isc.org. We have recently modified BIND9 DNS server with new DNS entries containing the IP addresses of proxy servers to be used in the indirect routes [100]. A secure DNS update program was developed to allow a coordinator to update DNS server with the new indirect routing entries.

Choi and Das proposed a novel scheme, called the Applicative Indirect Routing (AIR) [29], to control network traffic congestion and refine route availability by coping with unreliable links quickly. The reroute is based on the selection of direct shared neighbors between the two nodes connected by the disrupted link. It is very similar to the link restoration approach in traditional network restoration techniques [32, 31]. In a generic sense, the indirect routes can be used to reroute traffic around the affected area, either caused by a failed node/link or by intrusion traffic. It would be useful to extend the reroute mechanism based on multiple paths and to utilize nodes that are further away from a disrupted link similar to those path-based network restoration approaches.

To set up multiple path indirect routes, the first step is to select the desired proxy servers. Different proxy server selections may result in significantly different performance. Similar problems, like mirror server and cache server placement and selection problems, are topics gaining interests recent years [103, 83, 47, 54, 58]. Both mirror server and cache server are used to replicate web content to improve the user-perceived performance and reduce the over-all network traffic. We have developed heuristic algorithms to choose the best mirror sites for parallel download from multiple mirror sites [37]. The preliminary performance results on simulated network and real network are very promising.

The dispersity routing proposed by Maxemchuk [63] used channel sharing to reduce queuing delay in store-and-forward networks. This research was extended to virtual circuit networks to deal with long, bursty data sources [64], where both redundant and non-redundant dispersity routing techniques were described. A taxonomy of dispersity routing schemes was presented in [15]. Dispersity was achieved in the application layer in connection-oriented real-time communication [14]. Further work employed dispersity routing techniques for fault tolerance in realtime networks [16]. Three major parameters - dispersity, redundancy, and disjointness were proposed to describe and simulate these systems. A survey presented in [45] illustrated various strategies, such as packet-by-packet or string mode, to give dispersion in different network configurations. It also depicted several dispersion issues, such as partial disjointness, re-sequencing proce-

dures, and routing algorithms, which require further study. It is critical that the information on disjointness be made available to the upper layers to guide the selection of a subset of multiple paths. Split multiple path routing (SMR), proposed in citeLee01, focuses on building and maintaining maximally disjoint paths; however, the load is distributed only in two routes per session. Design of efficient route discovery protocols that carry information on disjointness is a challenging problem. Study of the impact of topology changes on disjointness properties of existing routes and the design of route update protocols to propagate disjointness property information is very important. It is also crucial to investigate the relationship between the range of dynamic link bandwidth adjustment and the sustain-ability of disjoint routes. Research results are needed to help guide engineering designs of multiple path *ad hoc* networks.

Group key management system provides group rekeying services to distribute new key when members join or leave [101]. Group key can be used to enhance the security of the indirect routing and the group communications for the coordination among the IDR agents.

Application of multiple path techniques in mobile ad hoc networks seems to be a natural extension since it should help reduce the impact of unreliable wireless links and constantly changing topology. On-demand multiple path routing schemes studied in [73] maintained multiple routes that were utilized only when the primary root failed. However, traffic was not distributed to more than one path. Multiple source routing (MSR) was investigated in [99] that utilized a weighted round-robin heuristic-based scheduling strategy among multiple paths in order to distribute the load, but no analytical modeling of its performance was provided. Vutukury and Garcia-Luna-Aceves presented a distance vector multiple path routing algorithm, called MDVA [98], which uses a set of loop-free invariants to prevent the count-to-infinity problem. In [75], the positive effect of alternate path routing (APR) on load balancing and end-to-end delay in mobile ad hoc networks was explored.

In [55], a per-packet allocation granularity for a multiple path source routing scheme was shown to perform better than a per-connection allocation, and its inherent capability to provide a large variety of services was pointed out. Chow, et al. [33] proposed traffic dispersion strategies for multimedia streaming and showed how to achieve high error-free frame rate based on the characteristics of the MPEG video structure. This work illustrated that a tight collaboration among network layers, in this case the network layer and the application layer, can achieve better performance.

We have developed preliminary code to set up indirect routes from clients via a proxy server to alternate gateway using IP tunnels. From the alternate gateway to the victim side server, we use plain IP packet. The resolver library on Linux was modified to accept the DNS query results with indirect routing information. The following table shows the comparison of ping performance of direct vs. indirect route in normal operation and under DDoS attack situations.

We have also investigated a framework for enhancing groupware security by integrating them group key management system. A secure group secure groupware for first responders was developed by integrated the Key-stone group rekeying server [101] with Jabber instant messaging system [39]. It provides secure basic text-based group chat, group file distribution, and remote display services. The secure groupware runs on Linux PDA or palmtop with MANET driver from NIST.

Proposed Work: We propose to improve the efficiency and reliability of the secure DNS protocol for supporting multiple indirect route based on our preliminary work that modified BIND9 and secure DNS update. We will submit IETF draft and share Internet community with our source code and simulation results. Since updating client DNS server with new routing information only affects future connection request that resolve DNS-IP address mapping using DNS query, immediately reroute of existing client connection can be done by setting up a server daemon on the legitimate client machine to listen to the reroute request from the

coordinator. The other less intrusive approach to inform the client proxy server about new indirect routes. We propose to design secure protocol for establishing such timely notification to clients and evaluate their trade-off. We will design versions of efficient algorithms to selecting subset of proxy servers for a specific user, a client subnet, or a group of subnets. Through simulation with large scale network model, the performance of these proxy server selection algorithms will be evaluated. Note that further attacks on the proxy servers in our secure collective internet defense architecture can be used to track the source of intrusion. We will take that into consideration when designing the enterprise IDS.

We propose to improve the reliability and exception handling of the group rekeying service and apply it to enhance the secure communications among the coordinators, proxy servers, and alternate gateways.

2.2 Enhance Transport with Multiple Path Indirect Routes

Previous Work: Chen et al. 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 [26]. 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, 20, 102] 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 [46] 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 loca-tion and the underlying network can help improve TCP performance. It will also be interesting to see how the ca-pability of the lower layer to sustain the stability of the link segments improves TCP performance in wireless net-works.

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 [97], 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. It is important to investigate how to distribute the data over multiple routes based on desirable

requirements such as security, performance, and reliability.

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.

2.3 Cooperative Intrusion Detection and Handling

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 SCOLD system.

Previous Work: 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 [91], GrIDS [27], EMERALD [79] and AAFID [93]. 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 [74] presents a decentralized method for autonomous but cooperative component systems to detect distributed attacks specified by signatures.

In [105] 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 [48] that remove redundant alarms and support the hu-man 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.

In [23], 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.

In [30], we have investigated how to deploy firewalls for protecting mobile ad hoc networks. We have developed a PEAP module for the freeRadius server and analysis the performance of PEAP and TTLS protocol performance for support secure wireless authentication.

Secure AODV protocol was proposed by Zapata et al in [104] 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 [49]. Perrig et al presents the SPIN protocol for wireless sensor networks [76]. Carmen, Kruus, and Matt analyzed the constraints and approaches for distributed sensor network security [21].

Proposed Work: 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 2.4, 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.

2.4 An Integrated Approach to 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 [18]. 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 [35, 36, 57] for representatives.

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 [25, 109]. In [109], 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, 17, 28, 56, 51, 107, 108, 110]. In [28], 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 [36] and adaptive WTP [57] for representatives. The algorithms can be tailored in achieving queueing-delay differentiation at the service side [28, 56]. 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 [107, 108, 110].

In this work, we want to propose innovative resource allocation and scheduling approaches which can

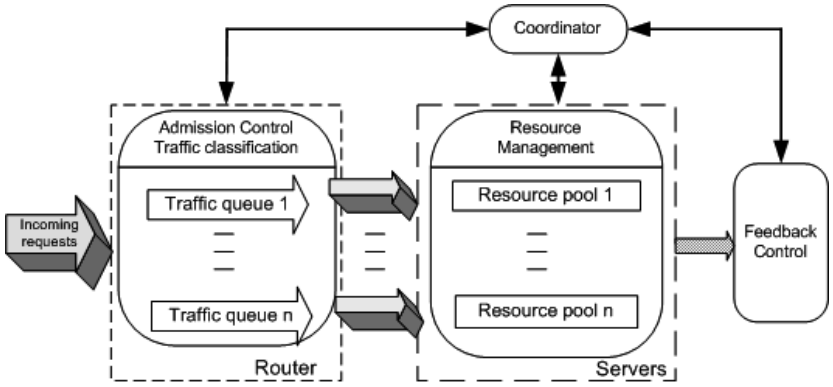

Figure 5: The architecture of the integration of admission control and adaptive resource management.

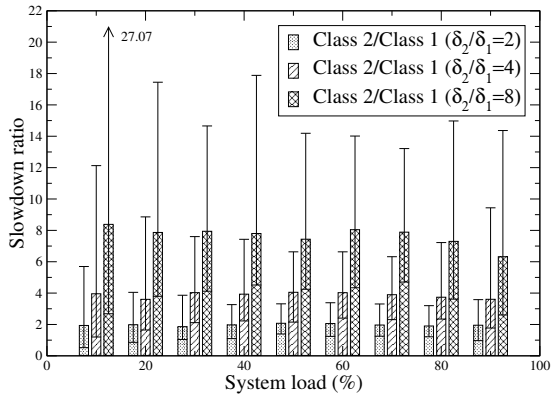

Figure 6: Percentiles of slowdown ratios.

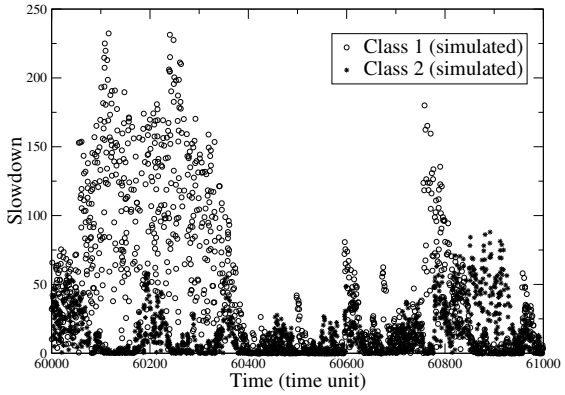

Figure 7: Slowdown of individual requests.

quantitatively control the QoS spacings between different traffic classes in terms of response time and slowdown. Figure 5 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 [108], 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 6 shows the percentiles of simulated slowdown ratios of two classes. The differentiation parameter ratios of class 2 to class 1 is 2, 4, and 8, respectively. 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. Figure 7 shows the simulated slowdowns of individual requests in short timescales when the server load is 90%. We can observe 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 [106], 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 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.

Figure 8 illustrates the implementation of the process allocation strategies on an Apache Web server. We adopted a two-class workload to evaluate the impact of the allocation strategies on the proportional response time differentiation. We implemented the two process allocation strategies by modifying child_main() function in http_main.c file of Apache. The process forking and killing mechanisms were not modified and still handled by Apache. This application-level implementation is hence flexible and portable.

Figure 9 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 9 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 [36, 57]. 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

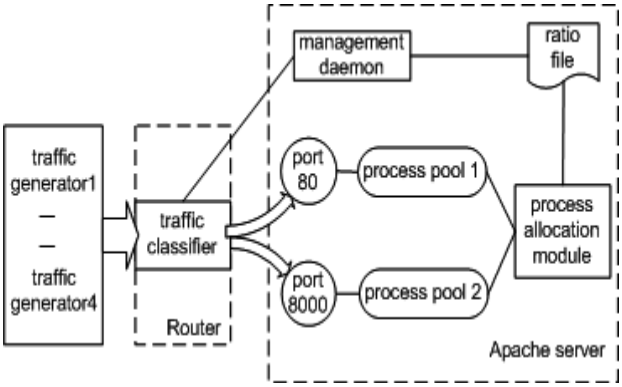

Figure 8: Implementation of process allocations.

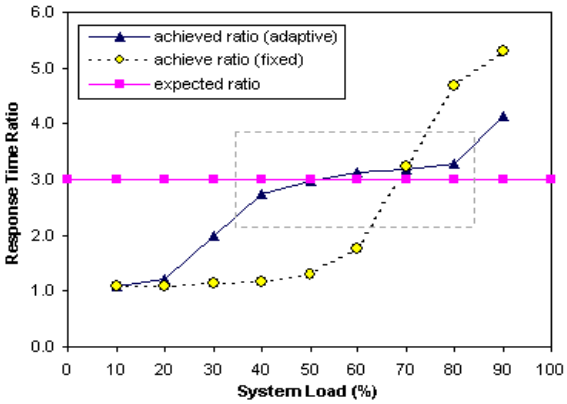

Figure 9: Impact of process allocations.

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 regulation incoming traffic under cyber threats.

2.5 Evaluation of Cyber-security Systems

In this section we address issues in evaluation of cyber-security systems with particular emphasis on predictability issues. This is inter-disciplinary work drawing on the experience of Dr. Boult in the areas of video-based physical security systems as well as his work in evaluation of biometric systems. The section begins with a discussion on the background of evaluation, reviews the state of the art, and then discusses the proposed work and timeline.

In the 70s and 80s, the term “intrusion detection system” was unknown by people and would have meant, to almost any “security” personnel, a system to detect the breach of their physical security perimeter. It may be been guards or growing array of electronic sensors along their fence-line and on their doors/windows. Long before we had significant computer networks, multi-sensor systems were monitoring the perimeter of sensitive installations looking for (physical) intrusions. To this day, physical intrusion-detection systems are an important aspect of physical security plans.

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 (CIDS) such as [82, 81, 4, 60, 59, 61, 65, 3, 24, 85, 96, 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), [43], 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.

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

However, as was discussed in [66], 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). 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[78]. 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 [71]. 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 [70]. 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 [89, 86, 53], 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 [42]. 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 [68, 69]. 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 [68], 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 [77, 44, 67].

There have been a few other "evaluation" metrics proposed. Stolfo, [95] 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, [40] 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 [96], 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 overlooked 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, [82, 81, 52, 34, 50, 60, 90]. 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, [66], 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 labeled, 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 [92]. While it would be good to proceed by extending the best existing protocols/software such as [85], they are, unfortunately, not available for public use.

More "realistic" data has been used in tests conducted by Nophais Labs, [88, 87, 72], 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[84], 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 [38, 72], we will

[2](though hinted at in [66]

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 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 [62], 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[80, 94, 13], but most cases the alarm is provided to a system administrator to manually address[22]. 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, [96] 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 [95] 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 randomizations 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

3 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).

In Fall of 2003 we held the first of these symposiums, with 4 speakers, a poster session and a Keynote address from E. Spafford. The symposium, sponsored by the Networking Information and Space Security Center, had attendance of over a hundred people, and included research presentations/posters from UCCS as well as 4 local institutions. To increase the outreach effort of this proposal, we propose (and have budgeted) to make the symposium a regular series.

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.

4 Research Schedule and Deliverables

We divide the execution of the proposal into three phrases and plan to conduct this project in three years:

1. The first year will be on
2. The second year will be on
3. The third year will be on

The research team will include the three PIs, four graduate research assistants, and three undergraduate research assistants. Dr. Chow has extensive experience in..... Dr. Boult has extensive experience in..... 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 .....Two GRAs will be working on..... One GRA will be working on..... One 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 servers, and 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 PI is affiliated.

The deliverables include ....., 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.

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] P. Albitz and C. Liu. *DNS and BIND*. O'Reilly & Associates, Inc., 2001.

[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] A. Banerjea. Simulation study of the capacity effects of dispersity routing for fault tolerant real-time channels. In *Proc. ECMAST*, 1996.

[15] A. Banerjea. Taxonomy of dispersity routing schemes for fault tolerant real-time channels. In *Proc. ECMAST*, 1996.

[16] A. Banerjea. On the use of dispersity routing for fault tolerant real-time channels. *European Transactions on Telecommunication*, 8(4):393–407, 1997.

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

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

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

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

[21] 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.

[22] 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.

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

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

[25] 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.

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

[27] 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.

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

[29] W. Choi and S. K. Das. Design and performance analysis of a proxy-based indirect routing scheme in ad hoc wireless networks. *Mobile Networks and Applications*, 8:499–515, 2003.

[30] 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.

[31] C. E. Chow and A. Hansmats. Design and analysis of one prong network restoration algorithms. In *Proc. of IASTED Conf. on Applied Informatic*, 1999.

[32] C.E. Chow, J. Bicknell, and S. Syed. Performance analysis of fast link restoration algorithms. In *Proc. of IASTED Conf. on Applied Informatic*, 1995.

[33] R. Chow, C. Lee, and J. Liu. Traffic dispersion strategies for multimedia streaming. In *Proc. of 8th IEEE Workshop on Future Trends on Distribute Computing*, 2001.

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

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

[36] 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.

[37] Chow E. and Y. Cai. Algorithms for selecting multiple mirror sites for parallel download. Technical report, UCCS CS technical Report, http://cs.uccs.edu/ scold/doc/mspd.doc, 2003.

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

[39] Jabber Software Foundation. Jabberx, jabber client. http://jabberx.jabberstudio.org.

[40] 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.

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

[42] Geof 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.

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

[44] 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.

[45] E. Gustafsson and G. Karlsson. A literature survey on traffic dispersion. *IEEE Network*, 11(2):28–36, 1997.

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

[47] S. Jamin, C. Jin, D. Raz, Y. Shavitt, and L. Zhang. On the placement of internet instrument. In *Proc. of IEEE INFOCOM, Mar. 2000*, 2000.

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

[49] 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.

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

[51] 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.

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

[53] 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.

[54] P. Krishnan, D. Raz, and Y. Shavitt. The cache location problem. *ACM/IEEE Transactions on Networking*, 8(5), 2000.

[55] R. Krishnan and J. A. Silvester. Choice of allocation granularity in multiple path source routing schemes. In *Proc. of IEEE INFOCOM*, 1993.

[56] 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.

[57] 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.

[58] B. Li, M. J. Golin, G. F. Ialiano, and X. Deng. On the optimal placement of web proxies in the internet. In *Proc. of IEEE INFOCOM*, 1999.

[59] 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.

[60] 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.

[61] 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.

[62] 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.

[63] N. F. Maxemchuk. Dispersity routing. In *Proc. of ICC*, 1975.

[64] N. F. Maxemchuk. Dispersity routing in high-speed networks. *Computer Networks and ISDN Systems*, 25:645–661, 1993.

[65] 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.

[66] John 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.

[67] 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.

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

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

[70] Ross J. Micheals and Terrance 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.

[71] Ross J. Micheals and Terrance 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.

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

[73] A. Nasipuri and S.R. Das. On-demand multiple path routing for mobile ad hoc networks. In *Proc. of 8th Int'l Conf. on Computer Communications and Networks (ICCCN)*, 1999.

[74] 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.

[75] M. R. et al. Pearlman. On the impact of alternate path routing for load balancing in mobile ad hoc networks. In *Proc. of MobiHOC*, 2000.

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

[77] 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.

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

[79] 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.

[80] 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.

[81] 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.

[82] 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.

[83] Li Qiu, V. N. Padmanabhan, and G. M. Voelker. On the placement of web server replicas. In *Proc. of IEEE INFOCOM*, 2001.

[84] 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.

[85] 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.

[86] Jun 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.

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

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

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

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

[91] 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.

[92] 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.

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

[94] 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.

[95] Sal Stolfo, Wei Fan, Wenke Lee, Andreas Prodromidis, and Phil 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.

[96] 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.

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

[98] S. Vutukury and J.J. Garcia-Luna-Aceves. Mdva: a distance-vector multipath routing protocol,. In *Proc. of IEEE INFOCOM*, 2001.

[99] L. Wang, L. Zhang, Y. Shu, and M. Dong. Multipath source routing in wireless ad hoc networks. In *Proc. of IEEE INFOCOM*, 2000.

[100] David B. Wilkinson. Enhanced secure dns: A defense against ddos attack. Technical report, Master Thesis, UCCS, 2003.

[101] C. K. Wong and S. S. Lam. Keystone: A group key management service. In *Proc. of Int'l Conf. on Telecommunication*, 2000.

[102] 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.

[103] E. Yilmaz and Y. Manzano. Surveying formal and practical approaches for optimal placement of replicas on the web. Technical report, http://websrv.cs.fsu.edu/research/reports/TR-020701.pdf, 2001.

[104] 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.

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

[106] X. Zhou, Y. Cai, G. K. Godavari, and C. E. Chow. An adaptive process allocation strategy for proportional responsiveness differentiation on Web servers. Technical report, CS Department and NISSC Center, University of Colorado at Colorado Springs, February 2004.

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

[108] 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 Symposium (IPDPS)*, April 2004.

[109] 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*, accepted, to appear.

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