SECURITY-ENHANCED QUALITY OF SERVICE (SQOS):[†] A NETWORK ANALYSIS

Pitipatana Sakarindr, Nirwan Ansari, Roberto Rojas-Cessa, and Symeon Papavassiliou
Advanced Network Laboratory, Department of Electrical and Computer Engineering
New Jersey Institute of Technology, University Heights, Newark, NJ 07102 USA
{ps6, nirwan.ansari, rrojas, papavassiliou}@njit.edu

ABSTRACT

Security and Quality-of-Service (QoS) issues have traditionally been considered separately, with different objectives and implementation architectures. No protocol has been designed and implemented so far to parameterize security as a QoS parameter. However, it has been noted recently that security and QoS are highly intertwined; security mechanisms may severely affect QoS mechanisms in terms of network performance and data confidentiality. In addition, users are not given the choices on which security services and mechanisms as well as which security level should be applied to their traffic. A security-enhanced quality of service-based (SQoS) network has been presented recently, with two major objectives. One objective is to offer the users elastic choices on the treatment of messages with appropriate security mechanisms with respect to their own QoS and budget requirements. Another objective is to facilitate interaction between security and QoS mechanisms in the most efficient manner by providing information to each other. In this paper, the network performance is investigated by defining the utility functions and maximizing the user's benefits subject to a set of constraints.

I. INTRODUCTION

Security and Quality-of-Service (QoS) systems have traditionally been considered separately with different objectives and implementation architectures. No protocol has been designed and implemented so far to parameterize security as a QoS parameter. However, it has been noted recently that security and QoS are highly intertwined; security mechanisms may severely affect QoS mechanisms in terms of network performance and data confidentiality. In addition, the users are not given the choices on which security services and mechanisms as well as which security level should be applied to their traffic. The Security-enhanced Quality of Service-based (SQoS) network, presented in [1], has two major objectives. First, the two systems have mutually dependent performances. Security mechanisms can actually be strengthened and enhanced with information obtained by the QoS system. On the other hand, malicious activities on the network such as Denial-of Service (DoS) attacks, Denial-of-QoS attacks [8], and the bandwidth stealth might diminish the QoS performance significantly. With a secure system, network QoS may still be guaranteed even under attacks. Furthermore, security systems can be implemented to assure users that when any traffic flow is attacked, QoS of remaining flows may be preserved. Second, compared with the choices of QoS classes, no flexible security services have been offered and the users have no option to select the appropriate security mechanisms for their traffic. To overcome these problems, we have proposed in [1] an architecture that attempts to achieve two major objectives: 1) to allow information sharing and cooperation between the security system and the QoS system, and 2) to offer users a broad variety of security mechanisms enabled on their traffic. This paper investigates the network performance through the architecture design proposed in [1]. The paper is organized as follows: Section II provides the background of the SQoS network, Section III presents the SQoS network analysis, and Section IV discusses the cooperation between the QoS and security systems, followed by the conclusions in Section V.

II. BACKGROUND OF THE SQOS NETWORK

The background concepts of the SQoS network are discussed in three sub-sections as follows.

A. ASSUMPTIONS AND NOTATIONS

The SQoS network makes the following assumptions in order to achieve its two objectives: the security system is an additional component, consisting of a number of processors and memory space, to be added into the existing QoS system; the *home* Autonomous System (AS), to which a sender subscribes, always negotiates successfully with other *away* ASs, to which the intermediate routers, including a receiver, belong to, such that these away ASs' routers attempt to honor the security services requested by the sender unless the mechanism to execute the service is unknown or there are insufficient available resources. There is also the assumption on the user behavior that users have an upper limit on the amount they can afford, and they will not pay more for higher

[†] This work has been supported in part by National Science Foundation under Grant Awards 0435250 and 0423305.

security-enhanced QoS regardless of the premium offered. The resources in the SQoS network simply refer to the queuing buffer, security-related memory, CPU power, and network bandwidth. We assume that each resource has a finite capacity and can be shared, either temporally (CPU cycles and network bandwidth) or spatially (memory space and buffers) [2].

The following notations are defined:

$y \in \{1,...,Y\}$: User index, which can take up to *Y* users in the SQoS network.

$d_j(.)$: The delay occurred at the j[th] intermediate edge router.

$C_j(.)$: The cost function to perform a service at node *j*.

D : The upper bound of delay for the data flow.

η : The upper limit of the affordable cost.

x^g, d^g, l^g: The data rate, delay, and loss rate of non security-related service *g* of the QoS system, respectively.

$\beta_d, \beta_l, \beta_x$: The user sensitivity to delay, loss rate, and data rate of the QoS system, respectively.

τ_u, τ_d : The user sensitivity to the service unavailability and delay of the security system, respectively.

$j \in \{1,...,J\}$: Index of intermediate edge routers along the path between the sender and the receiver, which can be up to *J* routers.

$n \in \{1,...,N\}$: Index of security-related services offered (or served), which can be up to *N* services, dependent on each intermediate router.

$m \in \{1,...,M\}$: Index of service degree offered (or served), which can be up to *M* degrees, and generally four degrees (representing high, medium, low, and none) in every router.

$g \in \{1,...,G\}$: Index of non-security-related services, which can be up to *G* services.

S_m^n : A security-related service *n* with service degree *m*.

B. QUALITY OF SERVICE DIMENSIONS

As addressed in [1], security is considered as another dimension of QoS parameters, in addition to data rate, packet loss rate, delay, and pricing. Consider a video-conferencing application between the army's field commanders and generals in the Pentagon that requires high security properties, which include user authentication, message authentication, user access control, an effective encryption key length, and high quality of services. The last property includes high video quality with low delay, at the minimum cost. The SQoS-supported network aims to serve concurrently both security purposes and QoS purposes, by choosing the appropriate security services, while maintaining other QoS preferences within proper bounds.

Since security is considered as one dimension of QoS, the QoS parameters are simply divided into two groups: security-related group and non security-related group. The security-related group has three parameters: processing rate, delay, and blocking rate, while the non security-related group has three major parameters: data rate, loss rate, and delay. The processing rate is the rate at which the Service Engine (SE) can completely execute all security services requested per packet (unit in packet per second) or per flow. The blocking rate is the rate at which the SE in the router cannot perform the requested service regarding to insufficient resources, and is equivalent to a ratio of blocked services to requested services.

C. THE SQOS NETWORK ARCHITECTURE

Ref. [1] presents the architecture of the SQoS network, in which users are offered various customizable security services and the security system cooperates with the QoS system, thus improving both the network and security performance and responding to the security need of an individual user. To implement the SQoS scheme, a security vector was proposed to determine a number of customizable Security Services (SSs) with choices of customizable Service Degrees (SDs). Let $SSV = \|SSV\|_j$ be a security service vector, consisting of *j* security service vector portions, where each vector portion is dedicated to every intermediate router. The SSV portions of each intermediate router are in the generic form of

$$\|SSV\| = \bigcup_{j=1}^{J} \|SSV\|_j \equiv \{(SS^1, SD_m), ..., (SS^N, SD_m), [X]\}_1, ,$$
$$..., \{(SS^1, SD_m), ..., (SS^N, SD_m), [X]\}_J$$

where *[X]* denotes other information that can be attached such as estimated cost, time and data length. The SQoS network has two communication phases, the probing phase and the data transmission phase as follows.

1. PROBING PHASE

From Figure 1, the probing phase begins when the user requests the connection to be established. The querying user sends out a probing packet, containing a single requested-SSV (rSSV) portion for all intermediate edge routers. The rSSV portion in the probing phase is denoted as $\|rSSV\| = \{(SS^1, SD_m), ..., (SS^N, SD_m), [data\_length]\}$.

Along with the requested services, the size of data sample is attached such that the intermediate routers can estimate delay and processing cost. Upon an arrival of the probing packet, each intermediate router authenticates the user's identity and verifies the user privileges from the probe,

which is actually the rSSV. Then, the router inspects every requested service from SS^1 to SS^N and their service degrees. Consequently, the result is recorded into an available-SSV (aSSV) portion, which is denoted as $\|aSSV\|_j = \{(SS^1, SD_m), ..., (SS^N, SD_m), [delay, time\_process, cost]_{estimated}\}_j$ and is then attached serially after the rSSV portion. Each aSSV portion associates with every intermediate router. The estimated processing time, delay, and cost are also recorded in the aSSV portion. The probing packet is repeatedly forwarded through the intermediate routers to the receiver, who replies to the querying user with an acknowledgement (ACK) packet. If there are *J* intermediate edge routers along the path, the ACK packet carries *J* available-SSV portions, one for each router, denoted as

$$\|aSSV\| = \bigcup_{j=1}^{J} \|aSSV\|_j \equiv \{(SS^1, SD_m), ..., (SS^N, SD_m), [delay, time\_process, cost]_{estimated}\}_1 , ..., \{(SS^1, SD_m), ..., (SS^N, SD_m), [delay, time\_process, cost]_{estimated}\}_J .$$

At the end of the probing phase, the querying user retrieves information from all aSSV portions carried in the ACK packet. Information in each portion includes a pair of service and degree offered by each router, the estimated delay, processing time, and cost with regard to the sample data length (in bytes). Then, both security-related and non security-related utility functions are used to evaluate and maximize the user's benefits subject to several constraints. The evaluation details are discussed in Section III. If the benefits are not satisfied, the connection request is discarded. Otherwise, the user proceeds into the data transmission phase

During the probing phase, if the requested security services or their corresponding service degrees are not offered by the routers, it is referred to as an unavailable service. This case occurs when the current AS does not recognize the requested service that the home AS agrees with the querying user or has neither sufficient resource to perform the requested services nor performs at the requested service degree. The querying user may abandon the connection or agrees on the service with a lowered degree.

2. DATA TRANSMISSION PHASE

Satisfied with the evaluation result, the user starts the data transmission phase during which the data flow is attached with security-related information and sent through the network. In other words, the rSSV portions, one for each intermediate router, are attached into each data flow. The rSSV portions in the data transmission phase are denoted as

$$\|rSSV\| = \bigcup_{j=1}^{J} \|rSSV\|_j \equiv \{(SS^1, SD_m), ..., (SS^N, SD_m)\}_1 , ..., \{(SS^1, SD_m), ..., (SS^N, SD_m)\}_J$$

Upon an arrival at each router, a router picks up its associated rSSV portion and executes the security services requested individually. The requested services may be rejected if the requested service is unknown to the AS, or if the service degree is downgraded from the one chosen by the querying user or if the service is entirely unavailable due to insufficient resources. After the security services were served, each router records the results by replacing the corresponding rSSV portion with the aSSV portion to report the querying user and, in some specified cases, to the AS's administrator. Upon an arrival of the data packet, the receiver may reply either immediately upon an arrival of a data packet or after a delay for several data packets with an ACK packet.

The querying user retrieves information from the ACK packet, containing all aSSV portions, denoted as

$$\|aSSV\| = \bigcup_{j=1}^{J} \|aSSV\|_j \equiv \{(SS^1, SD_m), ..., (SS^N, SD_m), [delay, time\_process, cost]_{served}\}_1 , ..., \{(SS^1, SD_m), ..., (SS^N, SD_m), [delay, time\_process, cost]_{served}\}_J .$$

The user evaluates whether the packet has received the records of served services and other QoS requirements, along with the total cost, which will be charged into the user's account. The service provider also records the network performance to improve future services.

During the data transmission phase, the service unavailability case occurs when the requested security services (or their corresponding service degrees) could not be executed by the routers due to insufficient resources for either the requested services or the requested service degrees. The connection might be discarded, or continued with the lowered degree as per user request in the Security Service Level Agreement (SSLA). The querying user, notified about this dissatisfactory result, can make a claim to the service provider.

In this paper, the impacts of service unavailability cases during the probing phase and the data transmission phase are not similar since any dissatisfaction existed in the data transmission phase is higher than in the probing phase. That is because the requested services and corresponding service degrees are altered while transferring data without an initial agreement with the querying user. This incident is analogous to the incident when an ongoing call is cut off from the handoff process. In practical, the difference among the unavailability cases occurred during two phases could be a factor used in the pricing model defined by the service provider.

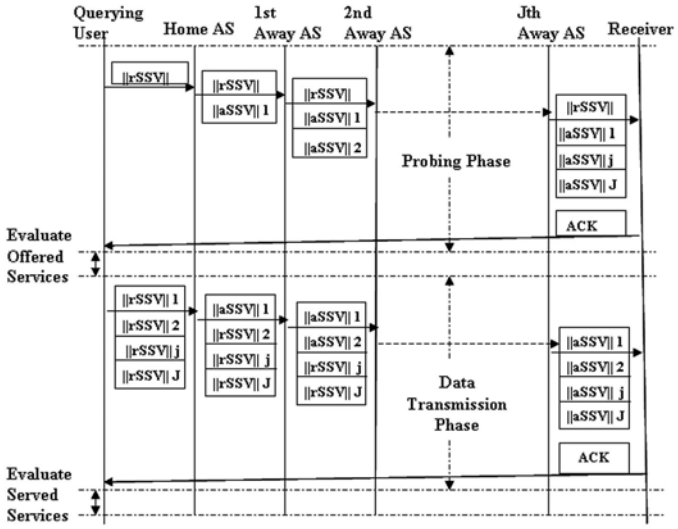

Figure 1. Transmission diagram of the probing and data transmission phases in the SQoS network.

For the sake of simplicity, the rejected service and downgraded service degree are considered as indistinguishable in this paper. Therefore, there are $J \times N$ security-related services offered. The details are discussed in the next section. Moreover, this paper does not take into consideration the significance of network dynamicity whereas the network status may change before the probing packet can reach back the sender, or whereas after the decision on connection establishment has been made with out-of-date information, retrieved from the probing packet. There is another issue concerning the network/link state update in QoS-based networks, but that is out of the scope of this paper.

III. A SQOS NETWORK ANALYSIS

The user determines a set of transmission parameters, including non security-related parameters (data rate, loss rate, and delay) and security-related parameters (processing rate, delay, and service and degree blocking rate). Objectively, these parameters should maximize the user's benefits, subject to the user's budget and transmission QoS requirements, as well as to optimize the network performance. The SQoS network adopts the Explicit Endpoint Admission Control (EEAC), proposed in [2], to configure a connection through intermediate routers and to determine:

1. whether there will be sufficient resources to guarantee all non security-related QoS requirements, and
2. whether there will be sufficient resources for executing the security services with their associated service degrees.

The user has the choice to terminate the connection request if the non-security-related parameters are not met, and if some of security services requested are rejected or performed at a downgraded service degree. By adopting the EEAC scheme, the SQoS network enables the end users, rather than the routers, to perform all necessary calculation tasks, which require powerful computation, and make a decision whether all available network resources can appropriately accommodate the user's both security-related and non security-related requirements.

A. A RESOURCE ALLOCATION MODEL AT THE SECURITY SYSTEM

As mentioned in Section II.C, we assume that a service i operated at degree $m = 1$ is independently different from a service i operated at degree $m = 2$. In other words, there will be $\mathbb{N} = \{1, ..., N \times M\}$ security-related services offered by each router. Let $\mathbb{R}$ be a finite set of feasible resource vectors in each router, $\mathbb{R} = \{R_1, ..., R_K\}$, and $R_k = \{R_1, ..., R_V\}$ denotes V portions of the k^{th} shared resource, and $R_k^{\max}$ denotes the maximum amount of the k^{th} shared resource. Therefore, the service i, $S^i; i \in \mathbb{N}$ is allocated with a portion of the k^{th} resource, represented by R_k^i, and all portions of the k^{th} resource must not exceed its upper bound such that $\sum_i^{\mathbb{N}} R_k^i \le R_k^{\max}$. In the SQoS network, the router resource is not reserved by the probing packet for the upcoming session. In the data transmission phase, the data packet carries the requested services, which are verified and complied by the service system.

A resource utility function that assigns a service i with resource R_k^i is referred to as $U(R_k^i)$. The resource utility function may be interpreted as a function of CPU power, memory, and bandwidth.

B. PROCESSING RATE ALLOCATION

From Figure 2, the Scheduling and Marking (S/M) module, as briefly illustrated in [1], classifies and marks the data packets to which class $q \in \{1, ..., Q\}$ they belong. A class is associated with the queuing parameter (δ). The S/M module provides the rate of incoming packets and their associated classes to the QoS Compliance Controller (QCC), where the packet that experiences the total delay exceeding the delay bound D will be dropped. The querying user is notified with the error message. The queuing status from the queuing system and the S/M module are also entered into QCC, yielding the output, which is a ratio between the processing rates among differentiated classes. The output is sent to the processing

rate controller to allocate the portions of service engine ($e \in \{1,...,E\}$) for performing the security services at the appropriate processing rate.

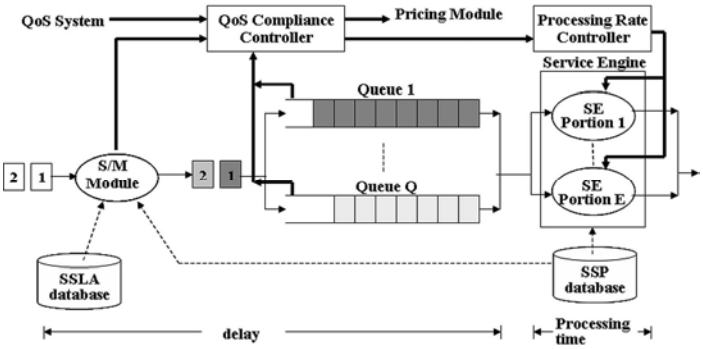

Figure 2. Service Engine and a Processing Rate Controller.

C. EVALUATION OF UTILITY FUNCTIONS IN THE SQOS NETWORK

Generally, a utility function determines a quantitative measure of the guaranteed QoS observed by the user [3], [4]. In the SQoS network, the user evaluates the security service vector twice at the end of the probing phase and data transmission phase. After the querying user receives an acknowledgement (ACK) packet in response to the probing packet during the probing phase (or the data flow during the data transmission phase) from the receiver, all security-related services offered (or served) by each intermediate router are evaluated to examine whether all SQoS requirements are satisfied. The cost of transmission and delay time will also be taken into consideration. To implement the scheme, both security-related and non security-related utility functions are evaluated with the objective to maximize the surplus of the utility function minus the cost function and the resource utility function, which is basically the user's benefit, subject to several bounds, including the cost, delay, resource, and capacity bounds. If the surplus is not found, the connection request is abandoned (or the services are blocked or served at the lowered degree).

The utility of services translates the value assigned by the user to the quality of both security-related and non security-related services. Thus, the output of the total utility function is defined as the sum of the output of the security-related utility function ($U_{security-related}$) and that of the non-security-related utility function ($U_{non\text{-}security-related}$), as follows:

$$U_{total} = U_{security-related} + U_{non-security-related},$$

where

$$U_{non-security-related} = \beta_x U\left(x^g\right) - \beta_d U\left(d^g\right) - \beta_l U\left(l^g\right),$$

$$U_{security-related} = U_p\left(S^i\right) - \tau_d U_d\left(S^i\right) - \tau_u U_u\left(S^i\right).$$

As mentioned earlier, there are three non-security-related utility functions in the SQoS network: data rate utility function $U\left(x^g\right)$, delay utility function $U\left(d^g\right)$, and loss rate utility function $U\left(l^g\right)$; and three security-related utility functions: processing rate utility function, $U_p\left(S_j^i\right)$, delay utility function, $U_d\left(S_j^i\right)$, and service unavailability utility function, $U_u\left(S_j^i\right)$. Following [3] and [5], the data rate utility function can be logarithmically proportional to the data transmission rate while the delay and loss rate utility functions can be linearly proportional to the delay and loss rate, respectively. For example, these functions can be defined as follows:

$$\beta_x U\left(x^g\right) = \beta_x \log\left(x^g / x_{\min}\right) + \Delta_x,$$

$$\beta_d U\left(d^g\right) = \beta_d d^g + \Delta_d,$$

$$\beta_l U\left(l^g\right) = \beta_l l^g + \Delta_l,$$

where $\Delta_x, \Delta_d,$ and Δ_l represent the constants used to adjust the offset of the data rate, delay, and loss rate utility functions, respectively. Since this paper focuses on the security issue, the validity of the non-security-related utility functions is beyond the scope.

For the security-related utility functions, the processing rate utility function $U_p\left(S^i\right)$ and delay utility function $U_d\left(S^i\right)$ are similar to the data rate utility function $U\left(x^g\right)$ and delay utility function $U\left(d^g\right)$ of the non-security-related functions, respectively. The service unavailability function may be calculated as a function of the blocking probability, defined as

$$U_u\left(S^i\right) = \frac{Number\ of\ unserved\ services}{Number\ of\ requested\ services}.$$

The objective of the evaluation of information retrieved from the ACK packet from the user's perspective is to maximize the user's benefit (the surplus), subject to a set of constraints from both the QoS and security systems, defined as follows:

$$\text{Maximize } \sum_i \left[U\left(S^i\right) - C\left(S^i\right) - U\left(R_k^{S^i}\right)\right]$$

subject to

1. cost bound

$$C_y = \sum_j^J \left\{\sum_j^{\mathbb{N}} \left[C_j\left(S^i\right)\right] + \sum_g^G \left[C_j\left(S^g\right)\right]\right\} \le \eta,$$

2. delay bound

$$\sum_{j}^{J}\left\{\sum_{i}^{\mathbb{N}}\left[d_j\left(S^i\right)\right]+\sum_{g}^{G}\left[d_j\left(S^g\right)\right]\right\}\le D \text{ , where}$$

$$\sum_{i}^{\mathbb{N}}\left[d_j\left(S^i\right)\right]>>\sum_{g}^{G}\left[d_j\left(S^g\right)\right],$$

3. resource bound

$$\sum_{i}^{\mathbb{N}}\left[R_k^{S^i}\right]\le R_k^{\max} \text{ and } \sum_{k}^{K}\sum_{i}^{\mathbb{N}}\left[R_k^{S^i}\right]\le\left(1/\delta^y\right)R^{\max},$$

where C_y is the total cost of the flow charged to the user *y*, $C_j\left(S^i\right)$ the cost function billed by node *j* of an application A^g, and δ^y is the proportional queuing parameter. The cost bound is used to prevent the sum of costs charged by the two systems (QoS and security systems) per each traffic flow from exceeding the cost upper limit. The delay bound limits the delay caused by two systems to the maximum delay that the data flow allows. The delay caused by the security system is expected to be much higher than that caused by the QoS system. Note that the delay occurred in any intermediate edge router generally includes both the processing delay, and the waiting-in-queue delay. The resource bound limits the upper bound for the k^{th} resource portions allocated to all services requested by user *y*, and also prevents the resource utilization of user *y* from taking up all available resources. The term $\left(1/\delta^y\right)R^{\max}$ indicates that the available resources are proportionally differentiated per class with regard to the queuing parameter, and allocated to serve the services requested by user *y*. The sum of capacities utilized by all traffic in link *z* must not exceed the link's maximum capacity.

Ultimately, from the provider's perspective, the objective of the SQoS network is to define and maximize the system utility. By using the user's utility functions, the utility of the whole network may be derived as the sum of all utilities perceived by all connections from all *Y* users. The system utility maximization is an NP-hard problem [5], [6], [7], in which several approximation algorithms have been proposed as solutions. Importantly, note that an exact utility function, $U(.)$, might not be accurately given; the service provider may define its own utility functions according to the business prospects.

IV. COOPERATION BETWEEN THE QOS AND SECURITY SYSTEMS

Another objective of the SQoS network is to make the cooperation between the two systems more effective. For example, when the QoS system receives a very large amount of incoming traffic from the same source or for the same destination, the QoS system alerts the security system to verify whether the network is being under attack, and concurrently keeps the log file of that traffic for future analysis. The security system can automatically allow some critical QoS information to be served with the highest security service degree.

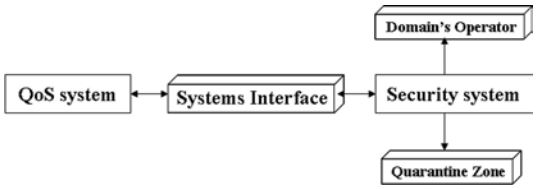

Figure 3. Cooperation between the QoS and security systems.

From Figure 3, the quarantine zone is deployed to keep some packets marked as suspicious as attacking packets while waiting for the user's confirmation. The cooperation between the QoS and security systems is classified into two types:

1. Active cooperation is referred as to when the network tries to prevent the QoS performance from deterioration caused by some specified attacks. Several researches, such as in [8], have addressed QoS attack scenarios, suggested detection methods, and proposed the solutions. In summary, several counter-attack schemes can be proposed based on this active cooperation.
2. Passive cooperation is referred as to when the victim from the attack and authorities attempt to regain information recorded during the attack and to trace the true identity of the attacker. The security system may also be integrated or cooperated with Intrusion Detection System (IDS) to detect ongoing attacks more efficiently. If there are suspicious activities, such as a large amount of traffic generated by a single source or destined to a single destination, the QoS system alerts the security system with in-depth information about involved traffic and the AS administrator is also informed. Early attack information can be recorded such that countermeasures or tracking processes can be rapidly performed.

The system interface translates the messages exchanged between the two systems to make messages understandable to each other.

Moreover, the AS administrator can configure directly the security system on-line by using router command packets, as illustrated in [1], in emergency cases, especially while being under attacks. For example, in a Distributed DoS (DDoS) attack scenario, the administrator may configure the edge routers to discard suspicious incoming traffic

destined to the victim's address and to quarantine others, which are waiting for authentication processing by the victim.

V. SUMMARY

Since users have different security requirements, they should be given the choices of security services. Consequently, the ISPs can assign more accurate resources and improve resource utilization. The SQoS network aims to achieve two major advantages: first, cooperation between security and QoS mechanisms to boost and secure the network performance; second, the SQoS network allows users to configure their preferred security services and service degrees for their traffic.

The network performance analysis is performed by optimizing the utility functions so that the result maximizes the user's benefits while maintaining the budget and satisfying the QoS requirements. System utility functions are presented to optimize the overall network performance. The utility maximization problem is proven to be NP-hard, and so the problem can be solved through approximation algorithms.

The cooperation between the QoS and security systems helps improve the overall network performance because some major network attacks could be prevented, and alleviated when the related systems share information.

REFERENCES

[1] P. Sakarindr, N. Ansari, R. Rojas-Cessa, and S. Papavassiliou, "Security-enhanced Quality of Service Design and Architecture," *Proc. IEEE Sarnoff Symposium on Advanced in Wired and Wireless Communications*, April 2005, pp. 129-132.

[2] J. Yang, J. Ye, S. Papavassiliou, and N. Ansari, "A flexible and distributed architecture for adaptive end-to-end QoS provisioning in next-generation networks," *IEEE Journal on Selected Areas in Communications*, Vol. 23, No. 2, February 2005, pp. 321-333.

[3] X. Wang and H. Schulzrinne, "Pricing Network Resources for Adaptive Applications in a Differentiated Services Network," *Proc. IEEE INFOCOM* 2001, Vol.2, April 2001, pp. 943-952.

[4] X. Wang and H. Schulzrinne, "An integrated resource negotiation, pricing, and QoS adaptation framework for multimedia applications," *IEEE Journal on Selected Areas in Communications*, Vol. 18, No. 12, December 2000, pp. 2514 –2529.

[5] T. O. Kamoto and T. Hayashi, "Analysis of service provider's profit by modeling customer's willingness to pay for IP QoS," *Proc. IEEE Global Telecommunications Conference*, 2002, Vol. 2, November 2002, pp. 1549-1553.

[6] C. Lee, J. Lehoczky, R. Rajkumar, and D. Siewiorek, "On quality of service optimization with discrete QoS options," *Proc. of Real-Time Technology and Applications Symposium*, 1999, June 1999, pp. 276-286.

[7] S. C. M. Lee, J. C. S. Lui, and D. K. Y. Yau, "A proportional-delay DiffServ-enabled Web server: admission control and dynamic adaptation," *IEEE Trans. on Parallel and Distributed Systems*, Vol. 15, No. 5, May 2004, pp. 385-400.

[8] W. Xiaoyong, V. A. Mahadik, and D. S. Reeves, "A Summary of Detection of Denial-of-QoS Attacks on DiffServ Networks," *Proc. DARPA Information Survivability Conference and Exposition, 2003*, Vol. 2, April 2003, pp. 277-282.