This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE Globecom 2010 proceedings.

Integrating Security with QoS in Next Generation Networks

Tarik Taleb*, Yassine Hadjadj Aoul#, and Abderrahim Benslimane+

*NEC Europe Ltd., Germany. #University of Rennes 1, France +University of Avignon, France.
{talebtarik, benslimane}@ieee.org, Yassine.Hadjadj-aoul@irisa.fr

***Abstract*—Along with recent Internet security threats, different security measures have emerged. Whilst these security schemes ensure a level of protection against security threats, they often have significant impact on the perceived Quality of Service (QoS). There is thus need to retrieve ways for an efficient integration of security requirements with their QoS counterparts.**

In this paper, we devise a Quality of Protection framework that tunes between security requirements and QoS using a multi-attribute decision making model. The performance of the proposed approach is evaluated and verified via a use-case study using computer simulations.

I. Introduction

Quality of Service (QoS) and security have been always addressed in a separate manner. However, studies indicate that security mechanisms require extra resources at both the network and end-users [2], often impacting the perceived QoS and sometimes degrading the overall system performance.

In the sphere of attempts to integrate QoS with security, researchers have recently started coining the term “Quality of Protection”, whereby sensitive information is protected using adequate authentication and cryptographic algorithms to ultimately ensure QoS [4]. For example, the work in [5] introduces a middleware adaptation scheme that dynamically tunes the encryption key length of the underlying encryption algorithm to the actual end-to-end delay. The major drawback of this work consists in its vulnerability to attacks such as man-in-the-middle [6]; in other words the work ensures a level of QoS but this comes at the price of some security flaws.

In this paper, we envision an “agile” framework, dubbed as QoS^2 (i.e., Quality of Service and Security), that protects the network from malicious usage and attacks. However, in the absence of a potential threat, the QoS^2 framework, in an autonomic way, relaxes the system’s overall security requirements in case the required QoS are not met under the current security settings. The framework provides an adjustable level of security to ensure acceptable QoS employing a Multi Attribute Decision Model (MADM) approach.

This paper is organized as follows. First, the relevance of this work to the state-of-art is presented in Section II. The envisioned QoS^2 framework is detailed in Section III. The performance evaluation of the proposed framework is presented in Section IV. The paper concludes in Section V.

II. Related work

In [1], Shen *et. al.* claimed that little work has been done on the interaction between QoS and Security in networks. They noticed that while QoS and Security used to be treated as separated entities, they strongly impact each other and thus should be considered together when designing protocols for Mobile Ad-Hoc Networks (MANETs). In [2], Irvine *et. al.* suggest a Quality of Security Service (QoSS) theory that handles security as a dimension of QoS. To enable service providers to advertise Security of Service (SoS) to their clients, researchers investigated the incorporation of security parameters into the Service Level Specifications (SLS) [7], [8]. The selected security parameters are integrated to enhance SLA-based management of QoS with the generation of network policies that guarantee the reservation of adequate resources for meeting both security and QoS needs.

To ensure both QoS and security requirements, some researchers addressed the problem using an adequate adaptive theory. For instance, the work in [9] exploits the co-operative game theory-based strategies to model the interaction between intruders and IDSs in MANETs and wired networks. Authors in [10] consider the tradeoff between security and resource consumption by formulating the problem as a nonzero-sum non-cooperative game between the certificate authority (CA) and an attacker. While these work, based on game theory and probabilistic models, may be able to formulate the problem of intrusion detection in a general case, they are not realistic for addressing the intricate interaction of multi-level QoS and security requirements. In order to formulate the QoS-security problem, we resort to the use of a distributed and elegant algorithm obtained from the Multi Attribute Decision Making (MADM) theory. MADM approach can be applied using different algorithms. Some of the commonly used ones are the Simple Additive Weighting (SAW) and Minimal Distance to Utopia Point (MDUP) algorithms. Authors in [13] conclude their performance evaluation that the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) algorithm outperforms both the SAW and MDUP algorithms. The advantages offered by MADM approach are more evident with the implementation of TOPSIS which we use in this paper.

III. Envisioned Solution

In the European research group ResumeNet [11][1], a taxonomy is developed to systematically document and assess the impact of various challenges, which pose a threat to the system. This taxonomy first identifies the challenge categories

[1] The present research work is part of the ResumeNet project.

978-1-4244-5638-3/10/$26.00 ©2010 IEEE

that reflect the nature of the challenge. A second-level classification is then formulated based on the specific scenario to which this challenge applies. The work then addresses the need to formulate defensive mechanisms for each of the challenges in different scenarios, and to also find the appropriate defensive measures. By performing rigorous system analysis and understanding the challenges that lead to the high likelihood of systems failures, in addition to learning from past events, in the ResumeNet project a library of the best possible defensive mechanisms specific to the given challenge and the prevailing scenario or set-up is built.

In this paper, we envision a "network security advisory system" with a number of threat levels ranging from low to severe. The security advisory system defines the threat level of the network based on events reported by entities such as firewalls and IDSs. It analyzes the events in specific timeslots and constantly updates the threat level. For each threat level and each associated security level, a particular defensive measure can be applied, following the taxonomy developed in [11].

Based on the alert level indicated by the security advisory system, we are interested in devising a security/QoS policy control that indicates the security level that should correspond to a desired QoS level. When under the indicated threat level, the security advisory system recommends a range of security levels (e.g., range of encryption/decryption key lengths, anomaly detection score, worm signature lengths, etc), we are interested in finding out the highest security level that should be selected in a way that the QoS requirements of users are not compromised. If for a particular security level, from within the recommended range, the QoS requirements of users cannot be satisfied (i.e., this can be inferred from a learning phase), the security/QoS policy control unit asks for security relaxation. In contrast, if the network is under potential attack and the security advisory system recommends the highest security level, the system has to stick to the recommended level although this decision may compromise the required QoS. QoS relaxation (e.g., transmission rate adaptation), thus, becomes mandatory in such a case.

A. Envisioned architecture

An example deployment scenario could be a carrier transport network administrated by a particular Network Operator (NO) and connecting a number of content/service providers (i.e., clients of the network operator) to their customers, located in different local networks. The overall network topology comprises a number of Monitoring Stubs (MSs), which are intelligently deployed over the core network next to strategic routers. These MSs form a hierarchical threat detection system, consisting of two layers, namely Local Security Monitors (LSMs) and Metropolitan Security Monitors (MSM). The LSMs gather information about behavioral anomaly of a particular local network and deliver them to the respective MSMs. Upon receiving information about a suspicious event from one of its LSMs, a MSM filters through its database of past attacks and evaluates if the threat is real. If the suspicious event is matched with one of the previous threats, the MSM notifies the security advisory. The security advisory contains a library of existing attacks and their counter-measures (i.e., following the taxonomy developed in [11]). The security advisory verifies if a similar threat originated from any other local networks administrated by another MSM (e.g., in case of rapid worm spread or DDoS traffic). The security advisory also consults its library of counter-measures (i.e., built following the taxonomy developed in [11]) to find the most appropriate mechanism to combat against the arising challenge. Then, the security advisory defines the threat level and relays the information pertaining to the threat detection and counter mechanism to all MSMs, which in turn, forward the information to all collaborating LSMs. When security measures are to be partially or fully enforced at the network elements (e.g. routers), the security advisory instructions are communicated by LSMs and MSMs to the respective network elements. The security advisory also notifies the service managers of the different service providers of details on the on-going threat and triggers them to take adequate measures to adapt their QoS demands to the new network dynamics. Instructions on QoS adaptation/relaxation is communicated either to servers, to end users, or to both when required.

B. MADM-based QoS-security level decision model

Let SM and PM denote the group of QoS metrics (e.g., bandwidth, delay, packet loss rate, etc) and the group of security metrics (e.g., encryption/decryption key length, timeliness, etc)[2], respectively. Let N_s and N_p denote the number of the QoS and security metrics, respectively. We assume that a user's Service Level Agreement (SLA) has S levels of QoS and P levels of security. When a user[3] connects to the network, the system advisory of QoS^2 must serve the user with a security level in such a manner to maintain good QoS requirements (i.e., QoS level). The suggested procedure is as follows.

1) Step 1: Defining all possible QoS level and security level combinations: For each QoS level, there are some requirements related to the values of its metrics that should be respected. These values should not be beyond (resp., below) a predefined threshold for cost metrics (resp., for benefit metrics). For example, for a given QoS level, the bandwidth (BW) should exceed a threshold (e.g., $BW_{th} \leq BW$) and the delay should be less than a threshold (i.e., $D \leq D_{th}$).

On the other hand, when we apply a security level, the values of the QoS metrics get negatively influenced in general. This means that these values cannot exceed some thresholds that we can measure by experiments. Consequently, some combinations will be impossible. Indeed, let QoS_{sup} denotes the highest QoS level and Sec_{sup} denotes the highest security level. Let's suppose that for QoS_{sup}, the bandwidth (BW) is ($BW_{high} \leq BW$) and that for Sec_{sup}, the bandwidth

[2] The work in [12] provides an extensive list of important security and QoS metrics.

[3] In this paper, the term "user" has a wider scope as it refers to a client of the network operator (e.g., content/service provider). Additionally, a connection does not necessarily refer to an end-to-end connection between a server and a client, but it does rather refer to the communication path between a content/service provider and a local network where some of its subscribers reside.

will be such as $BW < BW_{high}$. Thus, the combination (Sec_{sup}, QoS_{sup}) is impossible. Hence, the possible combinations (alternatives) of a QoS level and a security level will be limited. Let J denote the number of these alternatives $(J < N_s \cdot N_p)$. For each QoS level $s \in \{1, ..., S\}$, the value of the QoS metric sm $(sm \in SM)$ would be such as $sm_s \in [(sm_s)_{min}, (sm_s)_{max}]$ (i.e., sm_s: the value of the metric sm as defined in the QoS level s). On the other hand, for a security level $p \in \{1, ..., P\}$, the value of the QoS metric sm (sm_p) will be such as $sm_p \in [(sm_p)_{min}, (sm_p)_{max}]$.

As for the values of security metrics, they may have exact values (e.g., encryption key length) or may belong to an interval (e.g., timeliness) like all QoS metrics. The values of security metrics do not intervene in the definition of the alternatives. Now the selection of a combination of a QoS level $(s \in \{1, ..., S\})$ and a QoP level $(p \in \{1, ..., P\})$ will not be possible unless the following condition is satisfied:

$$\{[(sm_s)_{min}, (sm_s)_{max}] \cap [(sm_p)_{min}, (sm_p)_{max}] \neq \Phi\}$$

In the remainder of this paper, an alternative (a combination) is denoted as j (i.e., $j \in \{1, ..., J\}$).

2) Step 2: Defining the Decision Matrix (DM) for a user's connection c: Let $X^c_{jk}(t)$ be the value of metric k measured at time instant t for a connection c when the QoS-security combination j is used (i.e., $k \in \{1, ..., K\}$ being the index of the QoS-security metric). As we deal with different kinds of metrics (e.g., some expressed in kbps, others in seconds, etc), we normalize their values to be able to compare among them. Depending on the provided service and the user requirements, the QoS-Security metrics may not have the same importance. Thus, we assign each normalized attribute a weight such as the sum of all the weights is equal to one. Using the normalized values of the different metrics multiplied by their relative weight, we obtain the Decision Matrix for a given user's connection c at a time instant t. It should be noted that a decision D taken at a time instant t remains valid for a period ΔT^c_D during which the system gathers information from monitoring stubs and the collaborating networks.

3) Step 3: Applying a MADM algorithm to find the best alternatives among the available ones: In order to find the ideal alternatives or utopia points, we need to account for two types of QoS^2 parameters, namely "cost" metrics (e.g., bandwidth, delay, security level) and "benefit" metrics (e.g., throughput and fairness). The objective of formulating an utopian vector is to maximize the benefit while minimizing the cost as much as possible. In a reverse way, it is also possible to obtain the knowledge pertaining to the worst alternatives or nadir points in DM^c.

The TOPSIS selection algorithm [13] is derived from the MADM theory to extend these two contrasting utopia and nadir points to exploit the knowledge of both. We employ the TOPSIS selection algorithm for choosing the appropriate security level for ensuring the appropriate QoS requirements.

IV. Performance Evaluation

Having described details on our MADM-based QoS^2 approach, we now direct our focus to its performance evaluation using the Network Simulator (NS3) [14]. Given their strict QoS requirements, we consider IPTV streaming services. As a network threat, we envision the spread of Internet worms in a number of local networks whereby a number of subscribers to the IPTV service are located. As a counter measure, we adopt a signature-based worm detection approach as in our previous research work [3]. Along with a vast spread of the Internet worm, the security advisory recommends to MSMs and LSMs the filtering of inbound and outbound traffic using a generated signature with a particular length. The longer the signature length is, the longer the filtering-due delay becomes [3]. This intuitively impacts the end-to-end delay between the content servers and the end-clients. At the security advisory, six threat levels are defined to the above mentioned security counter measure; each threat level is characterized by $i)$ a range of worm signature substring length L_{worm} (i.e., principally responsible for additional delays at routers collocated with LSMs and MSMs) and $ii)$ a minimum number of signature substrings S_{min} that should exist in a traffic flow to generate an alarm.

In the simulations, we consider a network topology similar to that of Fig. 1 with video data streamed from a single content provider (i.e., Service Provider) to N_u subscribers located in local network 1. To avoid multicast scenarios, each subscriber is receiving a different video content over a dedicated session. At the content provider side, different video traces, encoded in MPEG-4, are used and servers use User Datagram Protocol (UDP) to provision IPTV service. To simulate network dynamics, we input some worm-affected background traffic along the path between Core Edge Routers (CERs) 1 and 2 simulating different Variable Bit Rate (VBR) UDP flows. The sending rate of each UDP flow is varying during the course of the simulation and is randomly chosen every $1s$ in a way that Core Network Edge routers are operating at loads exceeding 70% their full capacity. In the simulations, a noticeable increase or decrease in the background traffic rate triggers the security advisory to increase or decrease the threat level. With no specific purpose in mind, the aggregate propagation delays of links between Provider Edge Router and Customer Edge Router is set to $50ms$. Without any loss of generality, all links are given a capacity equal to $50Mbps$ (i.e., customer/provider edge links as well as core network links). In order to remove limitations due to small buffer sizes on the network congestion, buffers equal to the bandwidth-delay product of the end-to-end link are used. Due mostly to its simplicity and its wide usage in today's switches and routers, all simulated routers use Drop-Tail as their packet-discarding policy. The data packet size is fixed at $1380B$. The client side has a limited playback buffer length B, set to $100pkts$. The IPTV streams are characterized by an average streaming rate denoted as R_p and equal to $92pkts/s$ (i.e., 1.0Mbps). Simulations were all run for 900s; a duration long enough to ensure that the system has reached a consistent behavior. The presented results are averaged over the simulated N_u $(= 15)$ subscribers, averaged again on the total simulation runs (i.e., 7 times).

As mentioned earlier, for each threat level the security advisory recommends a range of parameters for signature

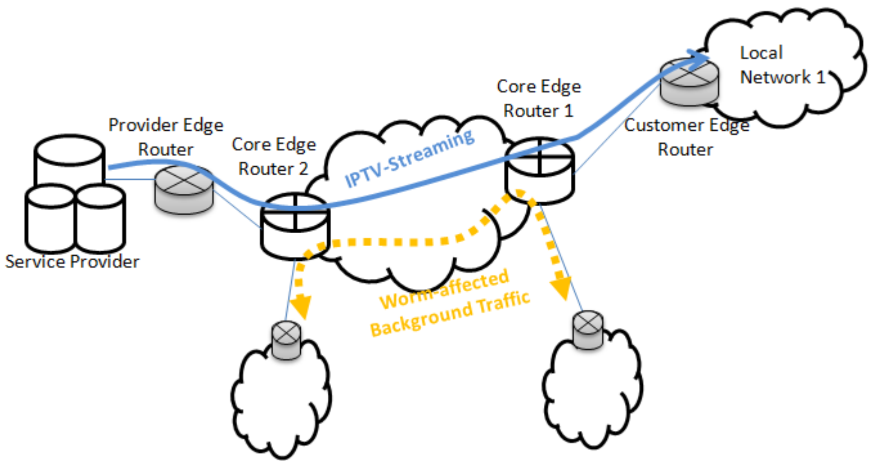

Fig. 1. Simulation network topology.

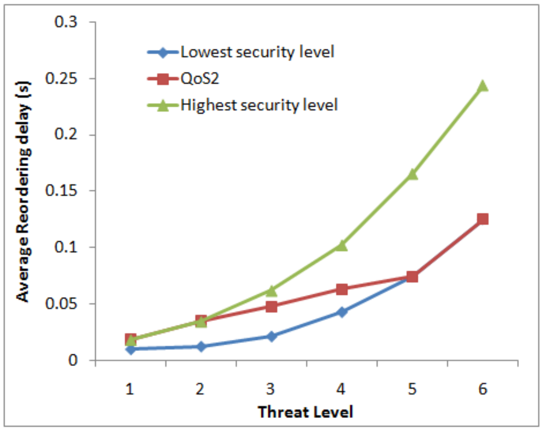

Fig. 2. Reordering delay for different threat levels.

generation, subsequent traffic filtering and worm detection. In the envisioned QoS^2 approach, the TOPSIS algorithm is run to select the best set of parameters (i.e., from within the recommended range) to meet both the QoS and security requirements. As comparison terms, we compare the performance of the QoS^2 approach against that of two conventional methods whereby the highest (i.e., longest) and the lowest (i.e., shortest) security levels (i.e., signature substring length) are selected. As parameters to quantify the users' perceived QoS, we consider the following metrics:

- Packet reordering delay: Difference between the arrival time of a packet and the arrival time of its preceding one. This metric is important for real time multimedia streaming services (e.g., IPTV) as if it exceeds the end-user's playback delay (i.e., $\frac{B}{R_p}$), the corresponding packet will be simply discarded before being transmitted to the application layer. The user shall then notice ruptures in the streaming service, a fact that impacts the perceived video quality.
- Playback ratio: Defined as ratio of the playback rate to the average streaming rate R_p. The playback rate is computed as the number of packets that were transmitted to the application layer every monitoring period of time (e.g., playback buffer delay $\frac{B}{R_p}$) and were indeed displayed over the monitoring period of time.
- Queue occupancy: The number of packets residing in the client's buffer and awaiting transmission to the application layer, measured every monitoring period of time (i.e., $\frac{B}{R_p}$). To achieve acceptable perceived QoS, this metric should not exceed the client's buffer size (i.e., overflow) and should not be in the vicinity of zero (i.e., underflow). Indeed, keeping a moderate value of this metric is highly important as it ensures for the application layer that there are always enough packets to display without having to discard them at the queue due to overflow.

Fig. 2 plots the average reordering delay experienced by the end-terminals for different threat levels and that is when the three security approaches (i.e., QoS^2 approach, and the security approaches applying the highest and the lowest security levels) are in use. Whilst the average reordering delay remains largely lower than the playback delay (i.e., $\frac{B}{R_p}$) for the three simulated approaches, in the simulations there were some instants when the value of this reordering delay exceeded the playback delay, and that is particularly when the highest security level is adopted in the event of high threat levels. The corresponding packets got intuitively discarded and this obviously would have been noticeable at the terminal's display. This will be manifested in poor playback ratio as indicated in Fig. 4. It should be noted that whilst the envisioned QoS^2 approach and the lowest security level exhibit relatively similar performance in Fig. 2, their main differentiator consists intuitively in their adopted security levels.

Figs. 3 and 4 show the impact of the three security approaches on QoS in terms of two correlating metrics, namely average buffer occupancy and playback rate, respectively. The two figures indicate an obvious observation: the best QoS is achieved when the lowest security level is adopted. However, this may not be acceptable from the security point of view. When the security measures are tight using parameters corresponding to the highest security level, the performance degrades significantly as there is not sufficient data at the terminal's buffer to play and consequently the playback rate is remarkably poor. In contract to these two approaches,

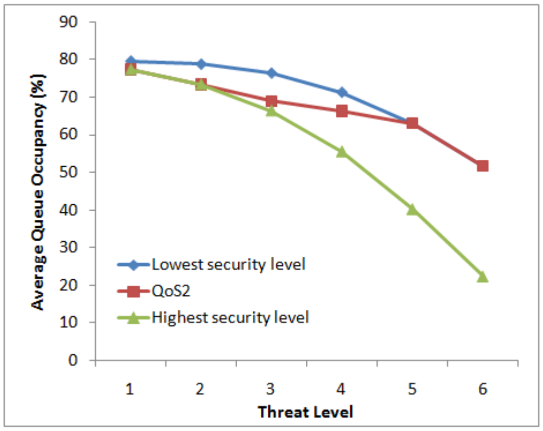

Fig. 3. Buffer occupancy for different threat levels.

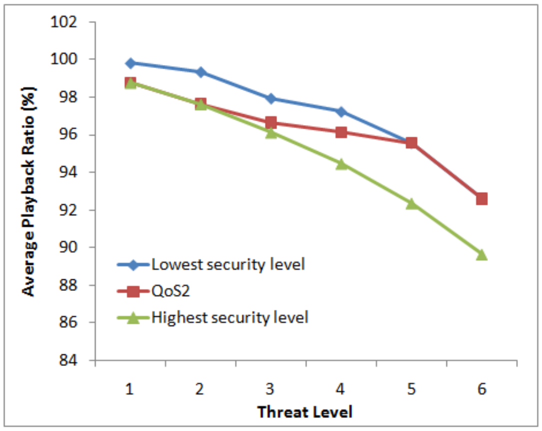

Fig. 4. Playback rate for different threat levels.

the proposed QoS^2 approach ensures an acceptable level of security and simultaneously achieves a QoS performance more or less similar to that obtained when the lowest security level is adopted.

The above simulation results show how the QoS^2 system jointly addresses the conflicting QoS and security requirements and demonstrate that adaptation of the security level according to QoS requirements yield satisfying results. However, it should be recalled that for a significantly high threat level, the advisory system may recommend significantly high security levels even if the QoS requirement are not met. In such events, QoS relaxation at servers and/or end-terminals becomes the only rescue.

V. Conclusion

In this paper, we demonstrated the need for jointly addressing QoS and security requirements. To this end, we devised a network policy framework entitled QoS^2 which orchestrates between the conflicting requirements of QoS and security based on a MADM approach running at a global security advisory system. The advisory system assesses current network security conditions based on real-time feedback from different monitoring systems deployed over the network in a hierarchical fashion.

We evaluated the performances of our QoS^2 mechanism while considering the case study of QoS-sensitive IPTV services. We demonstrate that our envisioned QoS^2 framework achieves its design goals. In the future, the proposed policy framework system is expected to evolve to cope with more complex network scenarios, different services, and more complex security counter measures that impact not only the E2E delay but also the bandwidth consumption and/or packet drops.

Acknowledgment

The work presented in this paper is supported by the European Commission, under Grant No. FP7-224619 (the ResumeNet project).

References

[1] Z. Shen, J.P. Thomas, "Security and QoS Self-Optimization in Mobile Ad Hoc Networks," in *IEEE Transactions on Mobile Computing*, Vol. 7 No. 9, September 2008, pp. 1138-1151.

[2] C. Irvine, T. Levin, E. Spyropoulou, and B. Allen, "Security as a dimension of quality of service in active service environments," in *Proc. Active Middleware Services Workshop*, San Francisco, CA, USA, Aug. 2001.

[3] K. Simkhada, T. Taleb, Y. Waizumi, A. Jamalipour, and Y. Nemoto, "Combating against internet worms in large-scale networks: an autonomic signature-based solution," *J. Security and Communication Networks*, Vol. 2, No. 1, pp. 11–28, Aug. 2008.

[4] S. N. Foley, S. Bistaerlli, B. O'Sullivan, J. Herbert, and G. Swart, "Multilevel security and quality of protection," in 1^{st} *Workshop on Quality of Protection*, Como, Italy, Sep. 2005.

[5] W. He and K. Nahrstedt, "An integrated solution to delay and security support in wireless networks," in *Proc. IEEE WCNC*, Las Vegas, NV, USA, Apr. 2006.

[6] Z. Fadlullah, T. Taleb, N. Nasser, and N. Kato, "Exploring the security requirements for quality of service in combined wired and wireless networks," in *Proc. IWCMC'09*, Leipzig, Germany, Jun. 2009.

[7] J. Jason, L. Rafalow, and E. Vyncke, "IPSec configuration policy information model," Network Working Group, *RFC 3585*, Aug. 2003.

[8] S. Duflos, V. C. Gay, B. Kervella, and E. Horlait, "Improving the SLA-based management of QoS for secure multimedia services," in *Proc.* 8^{th} *Int. Conf. on Management of Multimedia Networks and Services (MMNS'05)*, Barcelona, Spain, Oct. 2005.

[9] H. Otrok, M. Mehrandish, C. Assi, M. Debbabi, and P. Bhattacharya, "Game theoretic models for detecting network intrusions," *Computer Communications*, Vol. 31, No. 10, Jun. 2008. pp. 1934-1944.

[10] A. Rachedi, A. Benslimane, H. Otrok, N. Muhamed and M. Debbabi, "A Secure Mechanism Design-Based and Game Theoretical Model for MANETs," *Springer Journal of Mobile Networking and Applications (MONET)*, Vol. 15, No. 2, Apr 2010, pp. 191-204.

[11] M. Scholler and P. Smith, "Resilience and survivability for future networking: Framework, mechanisms, and experimental evaluation," *ResumeNet Deliverable D1.1b*, URL: www.resumenet.eu.

[12] G.A. Fink, B.L. Chappell, T.G. Turner, and K.F. O'Donoghue , "A metrics-based approach to intrusion detection system evaluation for distributed real-time systems," Pentagon Reports, Apr. 2002.

[13] F. Bari and V. Leung, "Multi-Attribute Network Selection by Iterative TOPSIS for Heterogeneous Wireless Access," in Proc. IEEE Consumer Communications and Networking Conference, Las Vegas, NV, USA, Jan. 2007.

[14] NS3, "The network simulator ns-3", available at URL: http://www.nsnam.org/, 2010.