HOMOGENEOUS SECURITY IN HETEROGENEOUS NETWORKS: TOWARDS A GENERIC SECURITY MANAGEMENT PROTOCOL

Justin P. Rohrer *
James P.G. Sterbenz
Information and Telecommunications Technology Center
The University of Kansas, Lawrence, KS, USA
Email: {*rohrej,jpgs*}*@ittc.ku.edu*
http://www.ittc.ku.edu/resilinets

Weichao Wang †
Dept. of SIS
Univ. of North Carolina at Charlotte, Charlotte, NC, USA
Email: *weichaowang@uncc.edu*

ABSTRACT

The promises of next-generation Internet architectures are many and varied. Key among these are increased flexibility and diversity. What this translates into is increased variety not only in protocols and technology, but in the performance characteristics and services provided by these composite or heterogeneous networks. This is certainly true of the diversity of emerging networking technologies being deployed in the current and next generation Internet environments. A significant concern is the need for a generic security protocol that is not tied to particular types of layer 1 and 2 networks. Based on lessons learned from the current Internet architecture we can see that this protocol needs to be able to communicate beyond trust boundaries (i.e. the borders of Autonomous Systems). We propose a high-level architecture for such a Generic Security Protocol and use simulation to demonstrate the benefits of having inter-domain communication in the context of the next-generation public Internet.

I. INTRODUCTION

The promises of next-generation Internet architectures are many and varied. Key among these are increased flexibility and diversity. What this translates into is increased variety not only in protocols and technology, but in the performance characteristics and services provided by these composite heterogeneous networks.

The Global Internet is increasingly incorporating diverse subnetwork technologies, motivating the need for accommodating heterogeneity. We are seeing ad-hoc and sensor networks with specialized multilayer protocol solutions, as well as networking technologies designed for use in space [9] and interplanetary [6] applications that add even more heterogeneity to the network in terms of long round trip times and highly asymmetric links. Along with the many benefits of enabling heterogeneous networks come numerous risks and challenges. One conspicuous issue is the need for a generic security protocol [11]. Previously proposed solutions to security concerns are generally designed to address particular scenarios and subnetwork technologies, and we believe that a more general solution is needed.

As the future of the Internet comes into focus we see that there is a tendency for policy, trust, and technology boundaries to coincide within the network topology. Since organizational units will determine both policy and implementation within their own boundaries, which may differ significantly from the choices made by bordering organizational units, this phenomenon is natural. However, it exacerbates a weakness that has already become apparent in the current, relatively homogeneous Internet, which is the lack of ability to effectively communicate policy beyond trust boundaries (namely past the borders of Autonomous Systems), described as tussle [8]. To counteract this, a generic security implementation of the next-generation Internet must explicitly provide mechanisms to facilitate policy dissemination across organizational boundaries.

In this paper, we focus on the problem of developing a generic security management service for heterogeneous networks, which will scale with the increasing diversity of network techniques and applications. This protocol will serve as a common language with which diverse heterogeneous network realms can exchange the appropriate security information and deploy generic and scalable security mechanisms. We consider security, survivability and resilience to be necessary features of all network components. We also believe that explicit provision needs to be made for the communication of security policy across trust boundaries. The term security can be used with varying scope in mind. For the purposes of this article we are defining security as those measures taken to ensure the health of the network as a whole and links individually.

* This work supported in part by NSF grant # CNS-0626949
† Work performed while at The University of Kansas

1-4244-1513-06/07/$25.00 ©2007 IEEE

II. RELATED WORK

Previous research on security issues in heterogeneous networks falls into three main categories, namely, authentication, collaboration incentives, and denial of service (DoS) prevention (summarized in [11]). The research on authentication [3, 17, 18] and collaboration incentives [12, 15] shares much in common with that targeted towards mobile and ad-hoc networks. DoS prevention falls into two categories: improved resource management [4, 2] and avoidance mechanisms [16, 10]. Improving resource management mechanisms is part of the natural evolution of technology over time. As it applies to low-capacity (eg. ad-hoc and sensor) networks this improvement can help reduce the disparity when they are connected to high-capacity networks. However, it must be realized that high-capacity network technologies (eg. wired and fiber-optic) are also continuing to be improved over time and generally at a faster pace so ultimately the principles of improved resource management lead to increased heterogeneity in the network and increased potential for DoS disruptions. This is because low capacity links can be saturated much more easily when high-capacity nodes are also present in the network. It is our belief that DoS prevention as well as remediation [14] must be addressed explicitly in the context of the next-generation Internet and our proposed solution is an attempt to do exactly that.

III. SECURITY CATEGORIES

As mentioned in the previous section, the security challenges driven by heterogeneous environments can be grouped under three main concepts [11], each of which need to be addressed for security to be possible. In this paper we are primarily concerned with the last of these three, namely DoS Prevention.

A. Authentication

Authentication is a necessary service for the establishment of trust. It can be handled in a distributed or centralized manner. Authentication paradigms have been rigorously studied in varied networking scenarios, some of which we perceive to be much more challenging than the next-generation Internet environment, which is expected to include trusted entities that lend themselves to the establishment of signed pubic key type authentication schemes. For this reason we do not intend to revisit authentication mechanisms for this protocol, but rather rely on those that already exist.

B. Collaboration Incentives

Incentives can be either a positive reward for correct behavior or a negative penalty for misbehavior. In the current Internet these incentives are generally negative and implemented through policy. Access to network resources is treated as a privilege which may be withdrawn if certain requirements are not met. While this is certainly not the only means available to ensure cooperation, there is a lot of momentum behind this mindset so for the purposes of this paper we will assume that future Internet architectures will operate on similar principles.

C. Denial of Service Prevention

Classically, DoS occurs as the result of intentional malicious behavior. In this paper we also recognize that DoS-like symptoms can result as a byproduct of the heterogeneity inherent in network designs due to congestion and saturation [19]. They are a possible disruption in any scenario in which a source or set of sources have the ability to generate more traffic than the lowest capacity link on the network path to an intended destination can support. The more the disparity of resources between different parts of the Internet, the more frequently such disruptions are likely to occur.

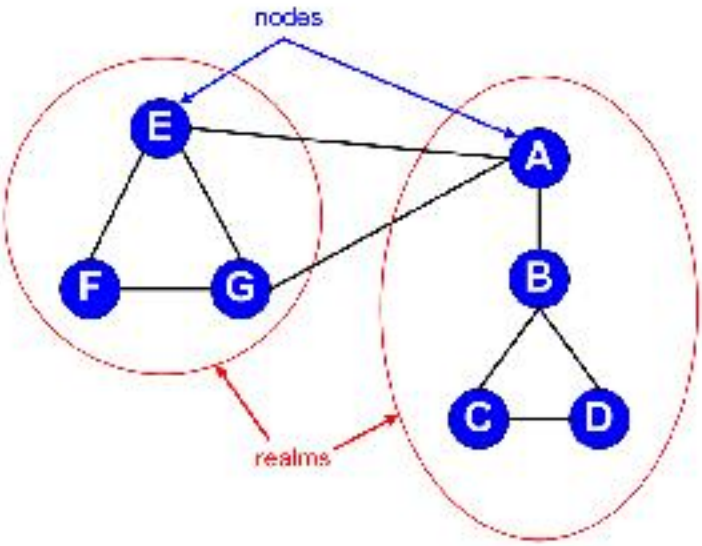

Figure 1. Next-Generation Internet Realms

IV. ENVIRONMENTAL MODEL AND GOALS

Emerging heterogeneous Internet architectures such as NewArch [7], the Postmodern Internetwork (PoMo) [5], and ANA [1] use the concept of realms (or compartments) to be collections of nodes that share mechanism and define trust and policy boundaries. A conceptual view of this model is shown in Figure 1.

We assume that there will be a core network infrastructure with its own governing authority similar to ICANN and IANA. We also assume that packets will be authenticated in such a way that their origin and path cannot be easily spoofed. Finally, we assume the need for cross-layer mechanisms to transfer policy information between end-users, administrative entities, and the network itself.

Our fundamental goal is to provide a generic security protocol to both unify policy implementation within each administrative realm and facilitate inter-realm policy communication to enhance the performance, resilience, and survivability of the network as a whole.

V. PROPOSED SOLUTION

Our Generic Security Protocol (GSP) architecture operates within the protocol stack influencing routing and forwarding decisions based on explicit policy configuration as well as input from outside entities such as firewalls. GSP relies on the services provided by the next generation internetwork and lower layers in order to function properly, yet is not an end-to-end protocol for users nor will it necessarily exist on all nodes in the network. The placement could be considered similar to that of BGP, although the functionality is entirely different.

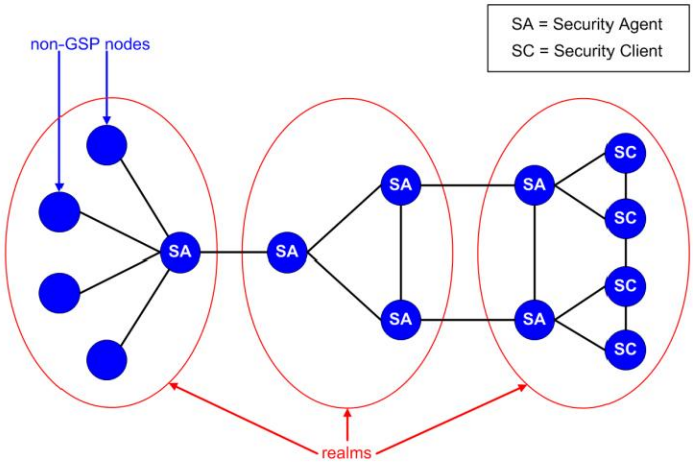

Figure 2. GSP Element Structure

A. *Framework*

Each realm will run a security agent service that may be implemented in a centralized or distributed manner within that realm. There are also optional Security Client services which run on individual nodes to allow them to receive policy information from the network in order to optimize their link usage. These entities share information using Security Packets. Figure 2 shows some possible configurations of security agents and Clients within different realms. In some realms it may not be necessary or practical to implement the GSP on all nodes. These realms will only need a single instance of a security agent to communicate with other realms. Other realms may need to run many security agents to distribute security policy to their realm peering points, or they may have nodes within the realm which can make use of security information and so run security clients on these nodes.

i. Security Agents

Security agents run on devices typically associated with the establishment or enforcement of security policies, i.e. gateways, firewalls, and intrusion prevention systems. Policy is defined around the basic link element and automatically distributed to all security agents within a given realm to enable cooperative enforcement. The security agent then interprets that policy to trigger a response when certain events occur. For example, if the firewall detects malicious traffic it can not only block the traffic itself but also signal the security agent with the flow ID and source of the traffic which will respond according to the local realm policy for that event type.

ii. Security Clients

Implementation of a security client on a given node is optional in that the decision is left to the end user or realm administration. The purpose of the Security Client is to receive relevant policy information from security agents within its realm and relay it upwards to the application and possibly the user.

iii. Security Packets

GSP signals using security packets. The security packet includes fields for *authentication*, *GSP code*, and *message*. The *authentication* field will authenticate both the source and the packet contents; *GSP codes* will be defined for specific types of packet; while *message* is an optional flexible field which could contain anything from an application specific command to a text directive such as "do not download copywrited music files". Security packets will be encapsulated in lower layer internetwork packets which will provide source, destination, and routing/forwarding information.

B. *Intra-Realm Security*

Within each realm, the primary concern of the GSP is the coordination of security measures and appliances to implement a unified security policy. This includes conveying security policy information from security agents to other security agents so that all security agents within a

given realm will have the latest policy updates, as well as conveying security event information from security agents to security clients.

C. Inter-Realm Security

One of the threats to network stability which we are addressing with the GSP is Denial of Service (DoS). Because GSP packets are a relatively small percentage of the network traffic and might be dropped in a DoS scenario, GSP traffic is prioritized above other types of traffic to ensure delivery.

D. Security for the Generic Security Protocol

With any type of automated policy enforcement comes the risk of exploitation through various types of attack against the new mechanism itself. We intend to mitigate this threat as much as possible through the use of public key encryption for authentication and strict requirements on GSP traffic behavior before it is allowed to enter the network or cross trust boundaries.

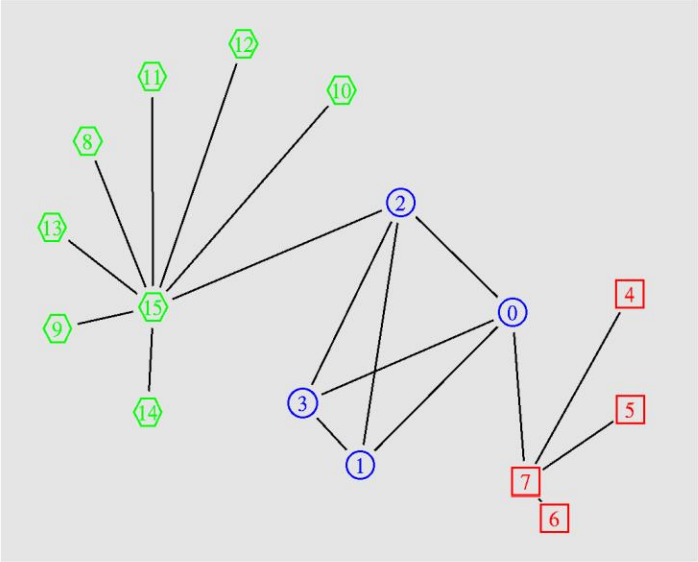

Figure 3. Simulation Model

VI. PROOF OF CONCEPT

While there are many scenarios that need to be tested and evaluated, we investigated one particular area in which inter-realm cooperation is essential for the prevention of network service degradation, namely traffic DoS challenges, whether resulting from deliberate attacks or unexpected but legitimate traffic beyond a link's intended capacity. To evaluate this we generated a simulation environment (shown in Figure 3) composed of three realms of different bandwidth capacities. The first realm (nodes 8–15) represents a set of end users or subscribers connected via low-bandwidth links. The second realm (nodes 0–3) represents a high-capacity provider realm. The third (nodes 4–7) contains a high-performance server farm. We then ran simulations to examine the disruption caused to well-behaving HTTP 1.1 traffic when misbehaving traffic is encountered on the low-capacity links both with and without the GSP.

A. Simulation Model

We created our simulation model using the open source ns-2 platform [13]. Each link was modeled using the symmetric duplex-link model and had droptail queuing implemented in both directions. The simulated legitimate traffic was generated using the built in PackMime HTTP-1.1 traffic generation agents with the default parameter set. Nodes 8–15 were set up as clients and nodes 4–7 as servers. Misbehaving traffic was simulated by creating a CBR connection between node 4 and node 14. This could represent many types of unwanted traffic, such as multimedia streaming or a malicious DoS attack. Traffic was monitored for a duration of 30 seconds on each run.

B. Results

Results were collected in the form of ns-2 trace files, and network animator (nam) format files, to allow both visual and numerical analysis of the simulations. Simulations were run to compare performance and response to disruptive traffic both with and without the GSP protocol operating. A control scenario without disruptive traffic was also evaluated.

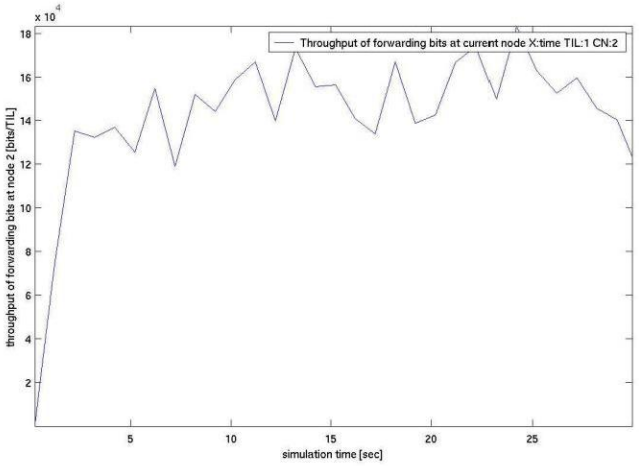

Figure 4. No CBR Traffic

i. Control Case

For the control case we ran the simulation with only well-behaved web traffic in the network. Successive iterations

of this simulation allowed us to tune the frequency of HTTP 1.1 requests such that only occasional packet loss was observed in the network as would be expected during normal use. This frequency was about one request per node per second. The flow of HTTP traffic was then maintained as a constant throughout the simulation in the other scenarios.

ii. Performance Without GSP

For the case without GSP operating, we defined node 15 to act as an ordinary firewall which detects and blocks the CBR traffic based on some rule set specified by the realm administrator. Because it has no way to signal upstream nodes there is little it can do to prevent the unwanted traffic from clogging it's downstream link from the service provider. In Figure 5 the upper curve represents the total traffic traversing the congested link, while the lower curve represents the well-behaved HTTP traffic. We can see that the goodput of the link drops dramatically when the CBR stream starts at time 17 sec. This is not only due to the packets being dropped due to congestion, but also because the TCP streams automatically back-off when they detect packet loss. We ran this with a variety of combinations of link speeds and CBR stream rates and found that the greater the disparity in resources, the more pronounced this effect becomes.

iii. Performance With GSP

In the case where GSP agents are running on all three realms we again set node 15 to detect and block the CBR traffic. Since all three realms are running security agents, node 15 is able to send signals upstream to node 7, which in turn can block the flow of unwanted traffic with minimal delay. If node 4 were running a Security Client node 7 would also be able to signal it to automatically stop the flow at its source, however this option was not included in our simulation. In this case (Figure 6) we can see a downward spike in goodput at the time the traffic disruption begins which rapidly returns to normal after the CBR flow is stopped. Comparing this to the graph from the control case (Figure 4) we can see that it is much more similar than the graph from the simulation without GSP.

C. Analysis

The idea of extending the communication of security policy beyond trust boundaries in the internetwork environment is a very powerful one. From this example we see dramatic improvement in the usability of the low-bandwidth network access link. The parameter values chosen (low bitrate = 128Kbps, high bitrate = 1Mbps, CBR bitrate = 320 Kbps, HTTP request freq. = 1/sec, GSP reaction time = 1 sec.) for the simulations are relatively conservative. Much larger bandwidth disparities already exist in the Internet and that differential will continue to grow. As they grow the benefit of implementing a protocol like GSP will also increase.

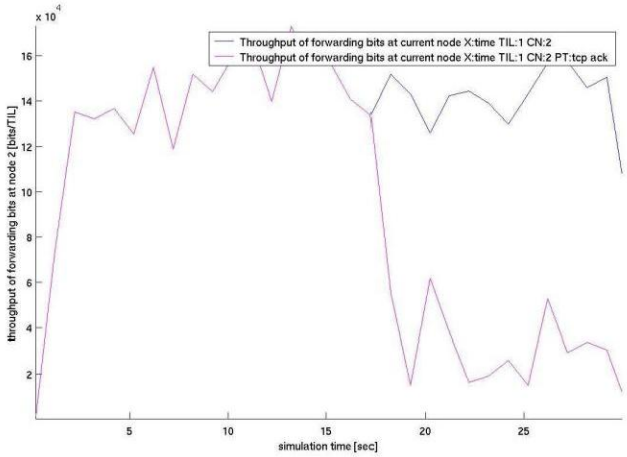

Figure 5. Without GSP

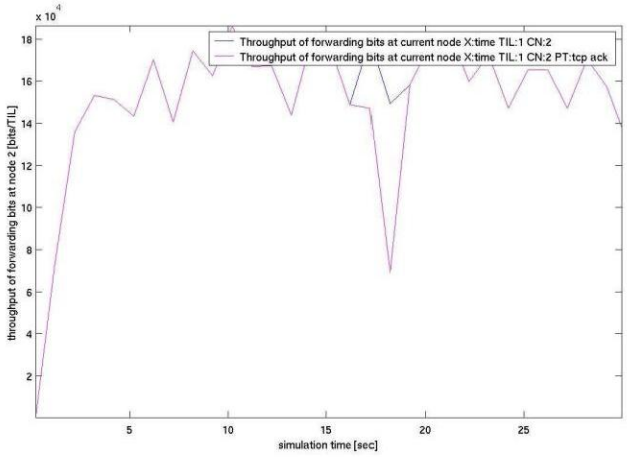

Figure 6. With GSP

VII. CONCLUSIONS AND FUTURE WORK

The simulation scenario presented here has demonstrated the potential power and flexibility of this approach. We believe that the same procedure could be used to prevent spam flooding or an SSH brute-force attacks as was applied to the DoS scenario in this case. It is not particularly difficult to see that stopping traffic that is disruptive to network operation close to the source is more desirable than blocking it close to the destination. The difficulty is overcoming the lack of trust between administrative domains, and also the issue of spoofed source addresses in the current Internet environment.

Next-generation Internet architectures have proposed solutions to overcome the latter issue, which should in turn go a long way to overcoming the former. Because of its reliance on these proposed solutions, published technical specifications for them are needed before our Generic Security Protocol can be defined in much more detail than what is presented in this article. Detection mechanisms for malicious traffic is a whole area of research beyond the scope of this paper. In the future we will investigate ways to integrate with existing traffic filtering solutions such as firewalls.

A. Security for the Generic Security Protocol

As mentioned previously it is imperative that any type of automated policy enforcement protocol such as this one be secured from the possibility of exploitation. We intend to thoroughly investigate the security measures necessary to achieve this using public key infrastructure, and traffic class management. We also intend to investigate the validity of using self-policing methods with compartmentalization for applications such as this one.

REFERENCES

[1] Autonomic network architecture: A European Union funded project in situated and autonomic communications, http://www.ana-project.org.

[2] S. Bali, Mitigating scheduler-induced starvation in 3g wireless networks, ITTC All-Hands Meeting, presentation (November 2006).

[3] B. Bhargava, X. Wu, Y. Lu, W. Wang, Integrating heterogeneous wireless technologies: a cellular aided mobile ad hoc network (cama), Mobile Networks and Applications 9 (4) (2004) 393–408.

[4] R. Bhatia, L. E. Li, H. Luo, R. Ramjee, P. Sanjoy, Icam: Integrated cellular and ad-hoc multicast, IEEE Transactions on Mobile Computing 5 (8) (2006) 1004–1015.

[5] B. Bhattacharjee, K. Calvert, J. Grifoen, N. Spring, J. Sterbenz, Postmodern internetwork architecture, Technical Report ITTC-FY2006-TR-45030-01, 2006.

[6] S. Burleigh, A. Hooke, L. Torgerson, K. Fall, V. Cerf, B. Durst, K. Scott, H. Weiss, Delay-tolerant networking: an approach to interplanetary internet, Communications Magazine, IEEE 41 (2003) 128–136.

[7] D. Clark, R. Braden, A. Falk, V. Pingali, Fara: reorganizing the addressing architecture, SIGCOMM Comput. Commun. Rev. 33 (4) (2003) 313–321.

[8] D. D. Clark, J. Wroclawski, K. R. Sollins, R. Braden, Tussle in cyberspace: defining tomorrow's internet, in: SIGCOMM '02: Proceedings of the 2002 conference on Applications, technologies, architectures, and protocols for computer communications, ACM Press, New York, NY, USA, 2002.

[9] R. C. Durst, G. J. Miller, E. J. Travis, Tcp extensions for space communications, in: MobiCom '96: Proceedings of the 2nd annual international conference on Mobile computing and networking, ACM Press, New York, NY, USA, 1996.

[10] W. Enck, P. Traynor, P. McDaniel, T. L. Porta, Exploiting open functionality in SMS-capable cellular networks, in: CCS '05: Proceedings of the 12th ACM conference on Computer and communications security, ACM Press, New York, NY, USA, 2005.

[11] J. B. Evans, W. Wang, B. J. Ewy, Wireless networking security: open issues in trust, management, interoperation and measurement, International Journal on Security and Networks 1 (1/2) (2006) 84–94.

[12] B. Lamparter, K. Paul, D. Westhoff, Charging support for ad hoc stub networks, Journal of Computer Communication 26 (13) (2003) 1504–1515.

[13] The network simulator: ns-2, http://www.isi.edu/nsnam/ns/.

[14] Resilinets strategy, http://wiki.ittc.ku.edu/resilinets wiki/ index.php/ResiliNets Strategy.

[15] N. B. Salem, L. Buttyan, J.-P. Hubaux, M. Jakobsson, A charging and rewarding scheme for packet forwarding in multi-hop cellular networks, in: MobiHoc '03: Proceedings of the 4th ACM international symposium on Mobile ad hoc networking & computing, ACM Press, New York, NY, USA, 2003.

[16] P. Traynor, W. Enck, P. McDaniel, T. L. Porta, Mitigating attacks on open functionality in sms-capable cellular networks, in: MobiCom '06: Proceedings of the 12th annual international conference on Mobile computing and networking, ACM Press, New York, NY, USA, 2006.

[17] W. Wang, W. Liang, A. K. Agarwal, Integration of authentication and mobility management in third generation and wlan data networks: Research articles, Wireless Communications and Mobile Computing 5 (6) (2005) 665–678.

[18] H. Yang, X. Meng, S. Lu, Self-organized network-layer security in mobile ad hoc networks, in: WiSE '02: Proceedings of the 3rd ACM workshop on Wireless security, ACM Press, New York, NY, USA, 2002.

[19] L. Xie, P. Smith, M. Banfield, H. Leopold, J. Sterbenz, D. Hutchison, Towards resilient networks using programmable networking technologies, Seventh Annual International Working Conference on Active and Programmable Networks (IWAN 2005), Sophia Antipolis, France, November 2005.