Available online at www.sciencedirect.com
ScienceDirect
Computer Networks 51 (2007) 3232–3251
Computer Networks
www.elsevier.com/locate/comnet

Modeling and performance evaluation of transport protocols for firewall control

Sebastian Kiesel *, Michael Scharf

Institute of Communication Networks and Computer Engineering, University of Stuttgart, Pfaffenwaldring 47, 70569 Stuttgart, Germany

Received 15 September 2006; accepted 20 November 2006
Available online 1 February 2007

Responsible Editor: I.F. Akyildiz

Abstract

Firewalls are a crucial building block for securing IP networks. The usage of out-of-band signaling protocols such as SIP for IP telephony and multimedia applications requires a dynamic control of these firewalls and imposes several challenges. Recently, several firewall control architectures and protocols have been developed. The main focus of this paper is the simple middlebox configuration protocol (SIMCO), which is a new transaction-based firewall control protocol. Due to the impact on call setup delays, firewall signaling requires small end-to-end delays and thus mandates a careful choice of the transport protocol. Therefore, this paper studies SCTP, TCP and UDP-based transport for SIMCO and compares different configurations that allow to optimize the performance. We present an analytical model to quantify the impact of head-of-line blocking in SCTP and TCP and verify it with measurements. Both the model and measurements reveal that SCTP can significantly reduce the SIMCO response times by leveraging transmission over multiple parallel streams. While already a few SCTP streams can almost completely avoid head-of-line blocking, our results show that TCP- and UDP-based transport may suffer from significantly larger delays.

Keywords: Firewall; IETF MIDCOM; IETF NSIS; SCTP; Head-of-line blocking

1. Introduction

Firewalls are a widely deployed technology to protect networks against unwanted access. The usage of out-of-band signaling in IP networks, namely "Voice over IP" (VoIP) solutions using the session initiation protocol (SIP), poses new challenges to firewalls. Due to the dynamic nature of SIP, firewalls have to take part in the session signaling. The advancement of firewall technology is further driven by the emerging IP telephony platforms, which are also referred to as next generation networks (NGN). These network architectures, such as 3GPP IMS or ETSI TISPAN, are intended to replace the circuit-switched telephone networks by

* Corresponding author. Tel.: +49 711 685 69017; fax: +49 711 685 67983.
E-mail addresses: kiesel@ikr.uni-stuttgart.de (S. Kiesel), scharf@ikr.uni-stuttgart.de (M. Scharf).

1389-1286/$ - see front matter
doi:10.1016/j.comnet.2006.11.031

SIP-based VoIP. However, compared to the Internet, these networks have much higher security requirements. Therefore, firewalls will be a major component in these platforms, e.g., for screening at the interconnection points of different operator's networks.

Several options exist for the signaling to such firewalls, such as NSIS or the simple middlebox configuration protocol (SIMCO), and there are many ongoing research and standardization activities in this field. Thus, the first part of this paper reviews different firewall control architectures and discusses their interaction with SIP-based applications.

Like most signaling applications, firewall control protocols have quite stringent delay requirements because the transaction delay contributes to the call setup delays perceived by users. Therefore, the choice and parametrization of transport layer protocols is of particular importance. The transmission control protocol (TCP) is the default choice for reliable transport in the Internet. Since TCP ensures reliable in-order delivery, end-to-end delays may be increased due to the head-of-line blocking effect when IP packets are lost. This effect is particularly critical on links with high data rates, i.e., between large softswitches and other central IP telephony platform entities such as firewalls.

While head-of-line blocking is a well-known problem of TCP, there are only few studies that quantify the impact of this effect on end-to-end delays. Potential alternatives to TCP for improving delays are multiple parallel TCP connections, multiple SCTP streams, SCTP unordered mode, and UDP-based transport. In the second part of this paper, we study the response time of transaction-based firewall signaling protocols over TCP and SCTP, both by analytical models and by measurements on different operating systems. This part extends work that has been published earlier in [1].

The remainder of this paper is organized as follows: in Section 2, we introduce fundamental firewall concepts and discuss the interaction of SIP signaling and firewalls. In Section 3 different architectures for firewall control are reviewed. Also, SIMCO as one promising signaling protocol for firewall control is presented. Sections 4 and 5 discuss the suitability of TCP, UDP and SCTP for signaling transport. At the example of SIMCO, we compare different approaches to reduce head-of-line blocking. Section 6 presents analytical models to quantify its impact. In Section 7, we present performance measurement results that have been obtained with a prototype implementation of "SIMCO over SCTP", and we compare the measurement results to our analytical models. Finally, Section 8 concludes this paper.

2. Securing IP telephony networks by firewalls

2.1. Firewalls in the Internet

With respect to computer networks, the term "firewall" is used to describe one or a group of network elements that enforce an access control policy on the traffic at the border between network domains with different security levels and requirements. That is, a firewall is basically a gateway that relays traffic from one domain to the other, but only if the traffic is compliant to a specific security policy. In the context of the Internet, firewalls are often used by organizations to protect computers *inside* a private network from unwanted access from the *outside* Internet [2].

To some extent, the concept of firewalls is contradictory to one of the design principles of the Internet, the end-to-end argument. However, firewalls are used because performing access control on incoming data flows at the receiving end system is often not sufficient: first, depending on the size of the local network it might be cumbersome to establish and maintain a consistent access control policy on all systems in the organization. Second, access control mechanisms on the end system might get compromised or disabled by uncooperative users, trojan horses, viruses, etc. Having a second line of defense in front of the end system might be a reasonable choice. While these two arguments are more of practical nature, the third one is also of architectural importance: an end system cannot defend itself against denial-of-service (DoS) attacks that try to disrupt the victim's connectivity by flooding its access link with useless data packets. Once the packet flood has traversed the congested access link, it is too late to filter these packets. Therefore, protection against this type of attacks has to be done before the bottleneck link, which is usually the link from the Internet service provider (ISP) to the local network, i.e., the firewall has to be placed at the edge of the core network.

Numerous scientific contributions deal with DoS attack prevention in the Internet. Usually, their focus is limited to protecting hosts that