Internet Censorship in China: Where Does the Filtering Occur?

Xueyang Xu, Z. Morley Mao, and J. Alex Halderman

Department of Computer Science and Engineering, University of Michigan,
2260 Hayward Street, Ann Arbor, MI 48109
{xueyang,zmao,jhalderm}@umich.edu
http://www.cse.umich.edu

Abstract. China filters Internet traffic in and out of the country. In order to circumvent the firewall, it is helpful to know where the filtering occurs. In this work, we explore the AS-level topology of China's network, and probe the firewall to find the locations of filtering devices. We find that even though most filtering occurs in border ASes, choke points also exist in many provincial networks. The result suggests that two major ISPs in China have different approaches placing filtering devices.

Keywords: Censorship, topology, network measurement.

1 Introduction

In this work, we explore where Intrusion Detection System (IDS) devices of the Great Firewall of China (GFC) are placed for keyword filtering at AS and router level. Knowing where IDSes are attached helps us better understand the infrastructure of the firewall, gain more knowledge about its behavior and find vantage point for future circumvention techniques.

China has the world's most complex Internet censorship system, featuring IP blocking, keyword filtering, DNS hijacking and so on [1]. IP blocking is the earliest filtering mechanism. It is easy to circumvent, because webmasters can always change their IP and DNS record. Besides, censors are very prudent to do DNS hijacking nowadays due to the risk of affecting the network in other countries [2]. In this paper, we focus on the most effective filtering mechanism of GFC, keyword filtering.

According to [4], the filtering occurs more at AS-level rather than strictly along the border routers. This paper answers the question whether all censorship occurs at border AS, and how filtering occurs inside those ASes. We first explore the AS-level topology of China's network. In this part, we explore which Chinese ASes are directly peered with foreign ones and which are internal ones. We call those peered with foreign network *border AS*, and the others *internal AS*. The resulting AS-level topology shows that the best vantage point to place filtering device is in the border ASes.

To find where IDS devices are attached at router level, we select a set of web servers in China and probe with HTTP GET packets that contain known

N. Spring and G. Riley (Eds.): PAM 2011, LNCS 6579, pp. 133–142, 2011.

keywords. In order to find more filtering devices, we manually select web servers to ensure their geographical diversity, as opposed to previous work that uses top websites in search result. This diversity is desirable, because it helps us to find more routing paths across China, and with more paths, we can discover more filtering devices.

The result shows that most filtering devices are in the border ASes, but a small portion is not. It is possible that there is a trend of placing filtering devices outside of border ASes. The number of router interfaces that have filtering devices attached for CHINANET is stable since 2007, while the second largest filtering force CNCGROUP has increasing number of filtering interfaces. Moreover, CHINANET's filtering is decentralized, while CNCGROUP has their IDS devices mostly in the backbone. A decentralized placement of filtering devices can facilitate censor to monitor domestic traffic.

The rest of the paper is organized as follows. Section 2 introduces the related work on measurement of the China's network censorship. Section 3 presents our result on AS topology of China's network. We locate filtering devices at router level in Section 4 to find how they are related to AS-level topology and the device placement strategies of different ISPs. Section 5 concludes the paper.

2 Related Work

An early work in the censorship measurement field is [3]. This paper analyzes the keyword filtering mechanism of GFC, and is a good source of background knowledge. They claim that the mechanism is based on an out-of-band intrusion detection system at border routers. The system emits forged reset packet to both destination and source, but packets themselves go through the router unhindered. Therefore, both source and destination ignoring forged reset packet makes the system entirely ineffective. They also claim that the firewall does not maintain a state.

An influential paper in this field is [4]. In the measurement study part, the most significant discovery is that unlike commonly believed, the censorship system in China is like a panopticon, where filtering does not occur strictly at border routers, but rather more centralized at AS level. They find that some filtering occurs 13 hops past border. In our work, we provide a more fine-grained analysis of where those filtering devices are located, answering whether all filtering occurs at border AS, and where IDS devices are attached at router level. They also discover that the firewall is stateful, namely a GET packet with keyword itself will not trigger the firewall. Rather, a complete TCP handshake is required. This contradicts with [3]. The paper also demonstrates that the RST packets sent by IDS devices are more complicated than before. The TTL of RST is now crafted, so we cannot identify the location of IDS devices by simply looking at the TTL values. Therefore, we identify the location of filtering devices by sending probe packets with increasing TTL values, and see when we receive RST packets from censors, as proposed in this work.

The most recent work in this field is [5]. This paper reports the discontinuation of keyword filtering in HTTP response on most routes, while that in HTTP GET