A Parallel Method in the 3G Firewall

ZhenYu Liu[1], ShengLi Xie[1], Yue Lai[1]

School of Electronic and Information Engineering,
South China University of Technology, Guangzhou, GuangDong, 510641, China
zy.liu@mail.scut.edu.cn

Abstract

IMSI (International Mobile Station Identity) is a unique number associated with all GSM and UMTS network mobile phone user. IMSI filtering—a prefix filtering, is an important function in the 3G firewall. It is an indefinite filtering which affects the efficiency of the security device. This paper brings forward a parallel processing method which uses vector coding and structure of bloom filter. This method is realized in Intel network processor—IXP 2850. The features of parallel multithreads and hash unit in network processor are taken advantages. The experiments show the excellent performance of the method in 3G firewall.

1. Introduction

IMSI is used to acquire details of the mobile in the Home Location Register (HLR) or as locally copied in the Visitor Location Register. In order to avoid the subscriber being identified and tracked by eavesdroppers on the radio interface, the IMSI is sent as rarely as possible. So IMSI is very important in the 3rd communication. [1-3]

IMSI filtering applies only to "Create PDP Request" messages. When performing IMSI filtering, the security device inspects GTP packets looking for IMSI that match IMSI rules. If the IMSI of a GTP packet matches an APN specified, the security device then refuses the package forwarding. But the IMSI filtering is an indefinite prefixal proceeding. The lengths of prefixes in IMSIs are not the same. This problem affects the efficiency of the firewall. [4]

The firewall is built in the Intel IXP2850 network processor — highest member of the Intel second-generation network processor product family. IXP2850 supplies a right platform for the 3G firewall in processing and encrypting high speed data between internet networks and 3rd Generation network.

Bloom filter is a space-efficient probabilistic data structure that is used to test whether an element is a member of a set. Some structures were brought forward to match strings [5-9]. A parallel method which takes advantages of network processor is shown. In this method, microengines provide support for software-controlled multi-threaded operations and the Hash Unit is used as a hash accelerator [10].

This paper is organized into five sections. The first section is a brief description of the paper. In section 2, the problems of 3G core network is provided and the 3G firewall is shown. The concepts of IMSI and IMSI filtering are introduced in section 3. In section 4, The parallel method is brought out. Vector coding, Bloom filter and IXP 2850 is introduced. And some experiments are used to exam the performance of the parallel the method. Section 5 is a conclusion of the paper.

2. 3G Firewall

2.1.3G Network

A 3G network includes two main sections: a Packet-Switched Core Network (CN), which is an IP-based backbone network, and one or more Radio Access Network (RAN). Along with the UMTS RAN (UTRAN) based on W-CDMA, Several operators maintain a parallel GPRS RAN evolved from the legacy GSM radio. This structure is sketched in Figure.1. [1-3]

The CN embeds several elements: SGSN (Serving GPRS Support Node), GGSN (GPRS Support Node), and a number of information servers. SGSN, being at the same hierarchical level as the Mobile Switching Center (MSC), keeps track of the location of a GPRS user, performs security functions, and handles access control. The GGSN is the logical gateway between the CN and external packet networks, such as Internet and

978-0-7695-3501-2/09 $25.00 © 2009 IEEE
DOI 10.1109/CMC.2009.255

private networks, is endowed with a full IP-stack and handles the IP-level connectivity with the MS.

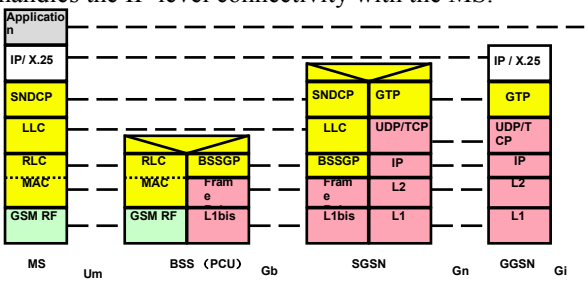

Figure.1 Structure of 3G Network

2.2. Security Problem

In 3rd Generation system, mobile stations can set up packet communications with terminal equipment connected to PDNs or another MSs in different PLMNs. To protect the privacy or user packet data, and to provide resualbe address capability in an intra-operator network, GTP is proposed.

GTP has no security to protect the communications between different 3rd Generation networks and the CN is based on IP. Attacks can originate from any interface, coming from external sources, such as the Internet or global roaming partners, or from within the provider's network, itself. These attacks may target the overall performance of the network and result in downtime, or exploit specific applications, such as the accounting and billing systems, and result in overbilling a customer for data services not used. Either way, a successful attack can result in real consequences, such as network degradation/downtime, lost revenue, disgruntled customers and possible customer attrition.

The security of any network is only as good as its weakest link, so operators must take precautions to secure their networks at all interfaces, particularly as they extend their circle of trust to include roaming partners. Obviously, the firewall is a good choice.

3. IMSI and IMSI Filtering

IMSI is an important field of GTP (GPRS Tunnelling Protocol) packet in 3G core network. The presentations of these are following. And the problem of IMSI filtering is analyzed.

3.1. IMSI

An IMSI is a unique number associated with all GSM and UMTS network mobile phone users. The IMSI is used in any mobile network that interconnects with other networks. It is stored in the SIM inside the phone and is sent by the phone to the network. It is also used to acquire other details of the mobile in the Home Location Register (HLR) or as locally copied in the Visitor Location Register. In order to avoid the subscriber being identified and tracked by eavesdroppers on the radio interface, the IMSI is sent as rarely as possible. [4]

The figure 2 is the structure of IMSI. An IMSI is usually 15 digits long, but can be shorter, for example MTN South Africa's old IMSIs that are still being used in the market are 14 digits. The first 3 digits are the Mobile Country Code, and are followed by the Mobile Network Code (MNC), either 2 digits (European standard) or 3 digits (North American standard). The remaining digits are the mobile subscriber identification number (MSIN) within the network's customer base. The MCC and MNC combined constitute the IMSI prefix and identify the mobile subscriber's home network, or Public Land Mobile Network (PLMN).

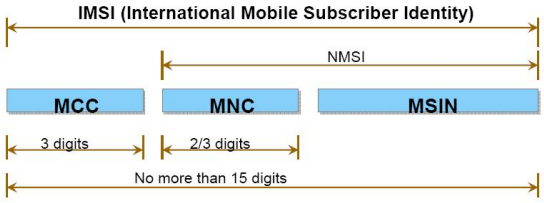

Figure2. The Structure of IMSI

3.2. IMSI Filtering

GTP filtering prevents some GSN device from "Create PDP Request" flood attack which consumes a great deal of storage and operations. If the IMSI of GTP packet matches the IMSI rules, the security device then refuses the package forwarding. So IMSI filtering is an important function in 3G security device. [4]

The security device is configured to deny GTP traffic coming from non-roaming partners by setting IMSI and IMSI prefixes. By setting IMSI and IMSI prefixes, the security device filters "create pdp request" messages and only permit GTP packets with IMSI prefixes that match the ones which is set. The security device drops GTP packets with IMSI prefixes that do not match any of the IMSI and IMSI prefixes that are set. Up to 1000 IMSI prefixes and IMSI can be set.

3.3. Problem of IMSI Filtering

IMSI filtering has two problems. One is from the structure of IMSI. Because the MNC is composed of 2 or 3 digits, so the IMSI prefix is indefinite. This will affect the proceeding efficiency.

Another problem is from the IMSI filtering which has two filtering－ IMSI prefix filtering and whole IMSI string filtering. For an example, a IMSI string "460011234567890" is to filtering, the prefix rules has "46001" and "460010" and the whole string rules has "460011234567890". When prefix filtering, the IMSI string is not matched. But when whole string filtering, it matches the rules. The problem is from the IMSI filtering which should match prefix and the whole string together.

4. Parallel Proceeding Method

A parallel method which uses prefix coding and a structure of bloom filter is brought forward. The realization of the method in network processor is also introduced.

4.1. Prefix Coding

Because an IMSI prefix is composed of 5 or 6 digits, the Trie algorithms need a great deal of space to search value from 0 to 0x99999 or 0x999999.

So a fast searching method which saves the space is needed. From the features of IMSI which is made up of digits, so the range of IMSI will be reduced after binary conversion. After converting, the searching space of prefix is 0x1869F or 0x F423F. If a bit vector is used to express this range, 5-digit prefix vector will use only 10^5 bit, about 12.5K bytes and 6- digit prefix vector will use 10^6 bit, about 125K bytes.

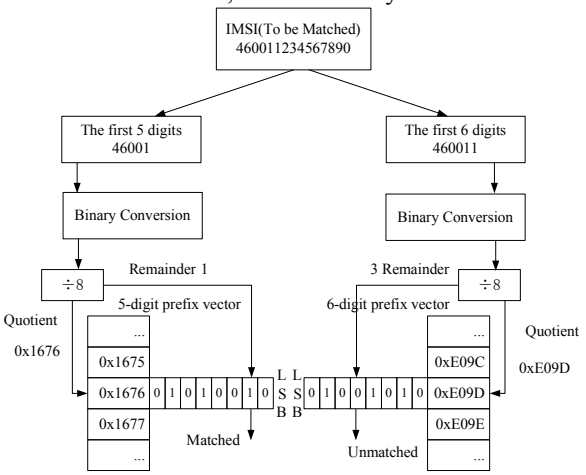

Figure 3 Searching of Prefix Coding

The searching method is shown in Figure 3. The first 5 or 6 digits of IMSI to be matched will be intercepted. The sub-strings of the IMSI are converted into binary digit. The binary sub-strings are divided by 8. The quotients find the bytes in the vector and the remainders locate bit in the bytes. In the study proceeding, the bit corresponding to the prefix is set to 1. In the matching, if the locating bit in the vector is 1 the IMSI is matched, otherwise unmatched.

4.2. Whole String Matching

In the whole string matching, a structure of bloom structure is used.

A Bloom filter for representing a set $S=\{x_1, x_2, \ldots, x_n\}$ of n elements is described by an array of m bit, initially set to 0. A Bloom filter used k independent hash functions $h_1, \ldots, h_k$ with range $\{1, \ldots, m\}$. These hash functions map each item in the universe to a random number uniform over the range $\{1, \ldots, m\}$ for mathematical convenience. For each element $x \in S$, the bits $h_i(x)$ are set to 1 for $1 \le i \le k$. A location can be set to 1 multiple times but only the first change has an effect.

To check if an item y is in S, we check whether all $h_i(y)$ are set to 1. If not, then clearly y is not a member of S. If all $h_i(y)$ are set to 1, then y is in S, although wrong with some probability. Hence a Bloom filter may yield a false positive, where it suggests that an element y is S even though it is not. For many applications, false positives may be acceptable as long as their probability is sufficiently small. [5-9]

The probability of a false positive for an element not in the set, or the false positive rate, can be calculated in a straightforward fashion, given assumption that hash functions are perfectly random. After all the elements of S are hashed into the Bloom filter, the probability that a specific bit is still 0 is

$$(1-\frac{1}{m})^{kn} \approx e^{-kn/m} \qquad (1)$$

Let $p = e^{-kn/m}$. Then the probability of false positive $f^{BF}(m,k,n)$ is then:

$$f^{BF}(m,k,n) = (1-(1-\frac{1}{m})^{kn})^k \approx (1-e^{-kn/m})^k = (1-p)^k \qquad (2)$$

Suppose given m and n and to optimize for the number of hash functions. There are two competing force: using more hash functions gives more chance to find a 0 bit for an element that is not a member of S, but using fewer hash functions increases the fraction of 0 bits in the array. The optimal number of hash functions that minimizes $f^{BF}(m,k,n)$ as a function of k is easily found by taking the derivative. More conveniently, note that $f^{BF}(m,k,n)$ equals $\exp(k\ln(1-e^{-k^{n/m}}))$. Let $g(k)=k\ln(1-e^{-kn/m})$. Minimizing the false positive rate $f^{BF}(m,k,n)$ is equivalent to minimizing $g(k)$ with respect to k.

$$\frac{dg(k)}{dk} = \ln(1-e^{\ln f_0 / k}) + \frac{kn}{m}\frac{e^{-kn/m}}{1-e^{-kn/m}} \quad (3)$$

When the derivative of equation (3) is 0, it has:

$$k_{\min} = (\ln 2)(\frac{m}{n}) \quad (4)$$

This is a global minimum. Alternatively, using $p=e^{-kn/m}$:

$$g(k) = -\frac{m}{n}\ln(p)\ln(1-p) \quad (5)$$

From which symmetry reveals that the minimum value for $g(k)$ occurs when $p=1/2$, or equivalently $k=\ln 2 \times (m/n)$. In this case the false positive rate $f^{BF}(m,k,n)$ is $(1/2)^k=(0.6185)^{(m/n)}$. In practice, of course, k must be an integer, and smaller k might be preferred since they reduce the amount of computation necessary.

For the carrier-grade devices, the requirement of accuracy rate is up to 99.9999%. So the $f^{BF}(m,k,n)$ should be less than 10^{-6}. Because the maximum quantity of IMSI rules is 1000, in the parameters {n, m, k}, the n=1000. It is necessary to determinate the quantity k of filters and the length m of bit vector. For convenience, the m is the power of 2. The false probability of the k=3 to 6 are researched. From Figure 4 shows the results.

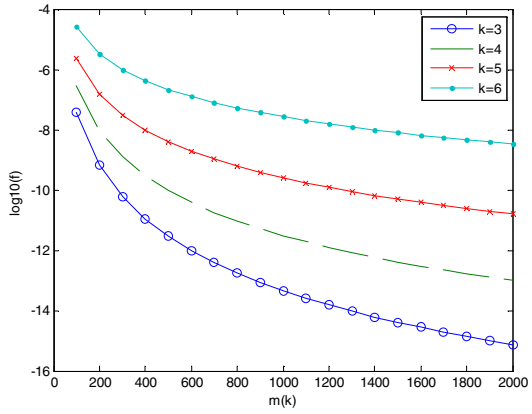

Figure 4. The False Probability of k=3 to 6 with m

From the Figure 4, when k=6, the $f^{BF}(m,k,n)$ is satisfied with the requirement of less 10^{-6} in any values of m. And the $f^{BF}(m,k,n)$ is the least in the all values of k.

4.3. Parallel Proceeding

4.3.1. Network Processor

Intel IXP2850 network processor delivers high-performance packet and content processing with robust security features in a single platform. The structure of IXP 2850 is shown in Figure 5. [10]

IXP2850 has 16 multi-threaded 1.4GHz microengines in the dataplane combined with a 700MHz Intel XScale™ core for control plane functions. The Microengine provides support for software-controlled, multi-threaded operation. There are eight hardware threads available in the Microengine.

In addition, the IXP2850 integrates a hash unit to accelerate the hash operation. The Hash Unit that can take 48-, 64-, or 128-bit data and produce a 48-, 64-, or a 128-bit hash index, respectively. The hashing algorithm in unit allows flexibility and uniqueness since it can be programmed to provide different results for a given input. The algorithm uses binary polynomial multiplication and division under modulo-2 addition. The operands are related by the equations:

$$A(x)\ M(x) = Q(x)\ G(x) + R(x) \quad (6)$$

In which, $A(x)$—input, M(x)—multiplier, $Q(x)$—quotient, $G(x)$—dividend, $R(x)$—remainder.

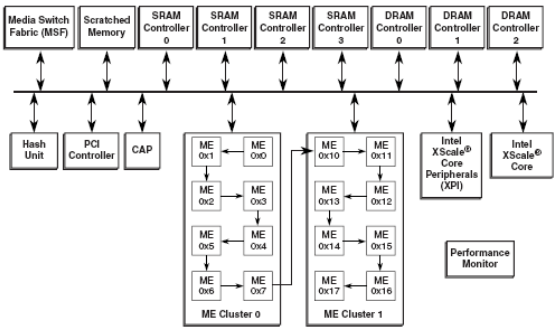

Figure 5. The Structure of IXP 2850

4.3.1. Realization

According to the structure of IXP 2850, the searching of vector with prefix coding and the whole string matching can be implemented together. The method is shown in Figure 5.

The 5 prefix matching and 6 prefix matching run in a thread respectively. Because eight hardware threads available in the one microengine, the prefix matching take two threads, so only six threads left to whole string matching. From the Figure 4, k=6 is enough to the requirement of less 10^{-6}. The thread which is used to whole string matching needs hash operation. Hash operation can be taken in the hash unit in network processor which accelerates the hash calculation. The hash function can be designed according to the requirement.

A match vector is used to decide whether the IMSI is matched to the rules. The first two least significant

bits judge whether the IMSI is matched to the prefix rules. The first bit and the second bit is corresponding to the results of 5 prefix matching and 6 prefix matching. Which one of these two bits are 1, the prefix is matched. If these bits are all 0, then prefix is failure.

The remains of the match vector are used to judge whether the whole string is matched. All these bits are 1 means the whole string is matched.

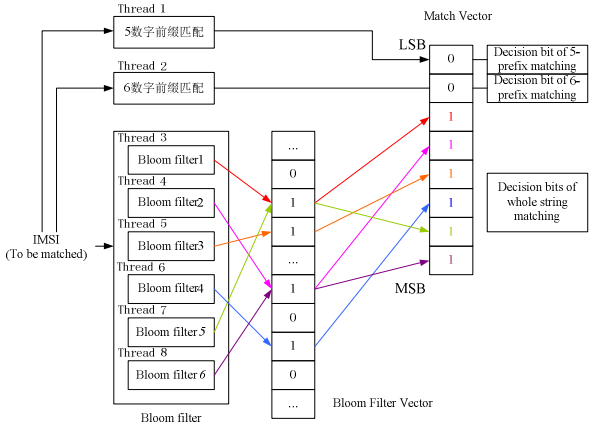

Figure 6. The Parallel Method

4.4. Experiment

The experiments investigate the performance of the parallel method. The quantity of IMSI rules in list is 1000. IMSI list are randomly generated. The testing IMSI are generated according to the IMSI rules. The testing IMSIs are edited in the Spirent TestCenter equipment and the throughout is tested in the case of k=1 to 6 and m=2000. The results are shown in Figure 7.

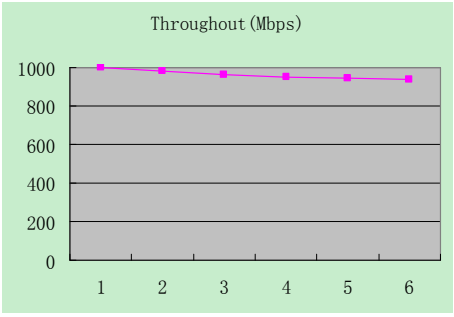

Figure 7. The Throughout of the Parallel with *k=1-6*

From the Figure 7, the throughout of the IMSI filtering is up to about 1Gbps. The throughout descends with the value of k increasing. When k=6, the performance is least. The main reason is the operations increase with the length of Bloom filter.

5. Conclusion

The paper introduces the 3G firewall and the important function—IMSI filtering. Through analysis of problems in IMSI filtering, a parallel method is brought forward. This method uses the vector coding and Bloom filter to match prefix and whole string together in network processor. The experiments show the good performance of this method.

Acknowledgment

The work is supported by National Natural Science Foundation of China for Excellent Youth (Grant 60325310), Guangdong Province Science Foundation for Program of Research Team (Grant 04205783).

References

[1] 3GPP, “GPRS Tunneling protocol (GTP) across the Gn and Gp interface”, TS29.060. v3.3.0, 2000.

[2] S. Li, C. Tsao, “Enhanced GTP: An Efficient Packet Tunneling Protocol for General Packet Radio Service”, *IEEE International Conference on Communications, ICC 2001*, Helsinki, pp. 2819-2823. 2001

[3] H.N.Hung, Y.B.Lin, “Connection failure detection mechanism of UMTS charging protocol,” *IEEE Transactions on Wireless Communication*, vol.5, NO.5, pp.1180-1186, 2006

[4] http://en.wikipedia.org/wiki/IMSI

[5] Juniper Corp. Juniper Networks NetScreen-500 GPRS.

[6] Bloom B. Space/Time trade-offs in Hash coding with allowable errors. Communications of the ACM, 1970, 13(7): pp.422-426.

[7] Andrei Broder, Michael Mitzenmacher. Network Applications of Bloom Filters: A Survey. Internet Math. Volume 1, Number 4 (2003), pp.485-509.

[8] Dharmapurikar S, Krishnamurthy P, Sproull T, Lockwood J. Deep packet inspection using parallel Bloom filters. In: Proc. of the Symp. On High Performance Interconnects (HotI). Stanford, 2003. pp.44-51.

[9] M. V. Ramakrishna. Practical performance of Bloom Filters and parallel free-text searching. Communications of the ACM, 32(10):pp1237-1239, October 1989.

[10] Intel Crop. “Intel IXP2850 Network Processor, Hardware Reference Manual”, January 2004.