A Comparison of String Search Algorithms for Deep Packet Inspection

Kieran Hunt

September 3, 2015

Abstract

Contemporary deep packet inspection systems often rely on custom hardware or entrenched ideas about the string search algorithms used. These algorithms have mathematically provable time or space complexities however not much is empirically known about their performance on real-world packet datasets. We felt that some string search algorithms could produce results that differed from their theoretical performance within the context of packet inspection. Furthermore, we sought to show that even algorithms with similar theoretical performances could produce differing practical results. Our approach was to reimplement a variety of the established string search algorithms and run them through a diverse set of tests with both real-world and constructed datasets. Our tests found that the Bloom filter was the fastest overall, Rabin-Karp was the most memory efficient and that the Nave algorithm was the slowest. Furthermore we found that, although the Bloom and Cuckoo filters have the same theoretical time complexity, the cuckoo filter was almost twice as slow as its counterpart. These findings help to show which algorithms perform best in practice and can help future algorithm designers to improve on the current approaches. In practice we show which algorithms a designer of a deep packet inspection system should consider implementing based on our findings.

1 Introduction

Deep packet inspection (DPI) forms part of the packet filtering that takes place on a computer network. For the general routing of a packet it is only necessary to look as far as the packet's headers. DPI takes the inspection further by closely examining the packet's payload data; its purpose is to detect data in a packet that is interesting to the network administrator. Generally, data that is interesting to the administrator is viruses, spam, incomplete or corrupted protocols, and obvious intrusions.

Deep packet inspection is very prevalent today. From Internet Service Providers (ISP) to Corporate Networks. Packets are inspected for a plethora of reasons.

Today, many approaches exist to perform deep packet inspection. The approaches can be split into two categories: hardware and software.

1.1 Subsection Heading Here

Write your subsection text here.

2 Algorithms

Write your conclusion here.

3 Testing Environment

Write your conclusion here.

4 Results

Write your conclusion here.

5 Analysis

Write your conclusion here.

6 Conclusion

Write your conclusion here.

References

[1] A. Aho and M. Corasick. Efficient string matching: An aid to bibliographic search. *Communications of the ACM*, 1975.

[2] R. Baeza-Yates and G. Gonnet. A new approach to text searching. *Communications of the ACM*, 1992.

[3] B. Bloom. Space/time trade-offs in hash coding with allowable errors. *Communications of the ACM*, 1970.

[4] R. Boyer and J. Moore. A fast string searching algorithm. *Communications of the ACM*, 1977.

[5] B. Fan, D. Andersen, M. Kaminsky, and M. Mitzenmacher. Cuckoo filter: Practically better than bloom. In *Proceedings of the 10th ACM International on Conference on Emerging Networking Experiments and Technologies*, 2014.

[6] N. Horspool. Practical fast searching in strings. *Software: Practice and Experience*, 1980.

[7] R. Karp and M. Rabin. Efficient randomized pattern-matching algorithms. *IBM Journal of Research and Development*, 1987.

[8] D. Knuth, J. Morris, and V. Pratt. Fast pattern matching in strings. *SIAM Journal on Computing*, 1977.