Efficient Packet Classification Based on Entropy

Michal Kajan
CESNET, z.s.p.o.
Zikova 4, 160 00 Praha, Czech Republic
kajan@liberouter.org

Jan Kořenek
Brno University of Technology
Božetěchova 2, 612 66 Brno, Czech Republic
korenek@fit.vutbr.cz

Motivation

- advance in Internet technologies
- increasing number of connected users
- Internet turned into a platform for business

Strong need to provide security and advanced networking management:

- basic filtering – firewalls
- traffic shaping
- traffic policing
- traffic accounting & billing
- lawful interception
- VPN networks & QoS
- …

Packet classification – basic principles

Classification comprises:

- predefined set of rules (with given priority)
- algorithm
- packets

The aim is to find the highest priority matching rule and make a decision about discarding or forwarding the incoming packet at the input of the classifier.

Demands on classification algorithm:

- high throughput – it is not possible to discard packets at the input of classifier
- low memory consumption – especially in hardware implementations
- support of multiple dimensions (packet header fields)
- support of high number of filtering rules

Related work

There are two different groups of algorithms to be most considered in recent literature and research:

- crossproduct based (decomposition) methods
- decision tree based (n-dimensional space division)

In the first group, classification is performed in several basic steps. For each field the longest prefix match is searched. Subsequently, individual results are combined and processed in order to find the matching rule. Combination and processing is upon a particular algorithm. Existing latest algorithms include DCFL [5], MSCA [1], PHCA [3].

Second group of algorithms generally creates a decision tree during the phase of preprocessing of the input rules. Since the classification occurs over several packet header fields and every field represents one dimension in the classification space, these algorithms divide n-dimensional space into subspaces.

Rules represent n-dimensional cubes and packet is a point in this n-dimensional space. The aim is to find the smallest cube containing the point. The most popular approaches are inspired by HiCuts [2] algorithm.

Based on these observations we can further analyse the character of the input rules and inspire to construct a new classification algorithm.

Classification based on entropy

Each rule can be considered as a sequence of bits. The amount of information is not analysed throughout each header field separately (thus not analysing particular dimension in classification space) but throughout each bit column.

Algorithm constructs a decision tree representing the space of all possible classification results:

- in each column a probability of given bit value is computed
- results are ordered
- tree is constructed according the probability value

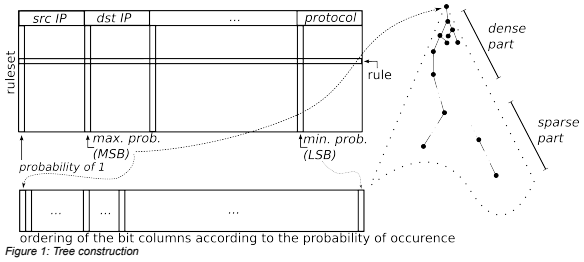

Figure 1: Tree construction

Higher parts of the tree close to the root node will be denser than the lower parts. Tree traversal from the root node to lower parts (reducing the search space) can be performed using a small number of bits in one step. On the other hand, in the lower parts hash functions processing high number of bits can be used to reach the leaf nodes and determine the matching rule. Further work involves development and testing methods for tree traversal.

Input rulesets analysis

Synthetic and real rulesets have been analysed – computation of probabilities (occurrence of bit value 1, 104 bits columns in total)

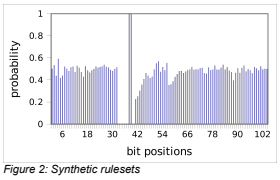

Figure 2: Synthetic rulesets

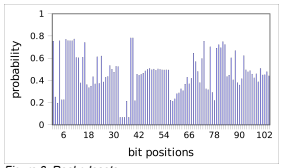

Figure 3: Real rulesets

Ruleset is represented as a table:

- row is one specific rule
- column is a bit of specific packet header field (src + dst IP, src + dst port, protocol)

References

[1] S. Dharmapurikar, H. Song, J. Turner, and J. Lockwood. Fast packet classification using Bloom filters. New York, NY, USA, 2006. ACM.

[2] P. Gupta and N. McKeown. Packet classification using hierarchical intelligent cuttings. In Proc. Hot Interconnects, 1999.

[3] V. Puš and J. Kořenek. Fast and scalable packet classification using perfect hash functions. In FPGA '09: Proceeding of the ACM/SIGDA, New York, NY, USA, 2009. ACM.

[4] S. Singh, F. Baboescu, G. Varghese, and J. Wang. Packet classification using multidimensional cutting, New York, NY, USA, 2003. ACM.

[5] D.Taylor and J. Turner. Scalable packet classification using distributed crossproducting of field labels., July 2005.

evropský sociální fond v ČR
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ