Hash functions

David Grochol

October 8, 2014

Abstract

The hash function is a mathematical function which converts an input data of the arbitrary size to the fixed size output data. The hash function requires a small difference in the input data to cause a large difference in the output. Hash functions have many applications eg. Hash tables that associate the output of the hash function and the corresponding data, another option is finding similar records, more frequent use is in cryptography, where the output is referred to as a checksum. The output of the hash function is called a hash value, hash code, hash sum, checksum or simply hash.

In order to apply the hash function, it must meet the following requirements (usually they are not all met completely): for any long entry, there must be a fixed output size, a small change in the input data will cause a large change in the output data. From the output of the hash function, it is computationally impossible to reconstruct the input data, the final requirement is that it must be highly unlikely for the two different inputs to correspond to the same output, it is used to verify whether the message has been corrupted or modified during transmission or not.

The aim of this work is to present a hash function mainly from a formal (mathematical) view, their characteristics and requirements. With the hash functions, the statistical distribution of the output in the whole range is also important. In this context, the evaluating methods of the quality of hash functions are introduced. Furthermore, also the perfect hashing and other ways of hashing are presented.

In the end, there are introduced the currently used hash functions. Their advantages, disadvantages, safety and time complexity of each functions.