Efficient SSD Integrity Verification Program Based on Combinatorial Group Theory

Xuejun Cai, Shenzhen Polytechnic, Shenzhen, China
Chaohai Xie, Shenzhen Polytechnic, Shenzhen, China
Xiaojun Wen, Shenzhen Polytechnic, Shenzhen, China

ABSTRACT

In digital forensics, the issue of data integrity protection for increasingly widespread applied SSD (Solid State Disk, SSD) is to be resolved. Based on Combinatorial Group Theory, mapping data objects in SSD validation process and test object in combination group testing methods, using the non-adaptive mode to the initial calculation, stored procedures, and re-calculate, verify process, and carefully selecting the design parameters to construct the test matrix of Combinatorial Group Testing method in response to different application environments, we design testing group. The algorithm capabilities meet the requirements and that is higher efficiently by repeated tests. It is feasible for SSD integrity check based on the Combinatorial Group Theory.

Keywords: Combinatorial Group Theory, Data Integrity, Map, Non-Adaptive, Test Matrix

INTRODUCTION

Computer Forensics is the process of providing legally binding electronic evidence for the entire forensic process via use of computers and related scientific and technical principles and methods. Computer evidence could be vulnerability, perishable, concealment and multi-media and other characteristics, moreover as time goes by, the computer evidence may change or disappear, so ensuring all the evidence collected from the original track and provided in court all keep it reality and integrity is most significant for improving the computer evidence’s probative value (DFRWS, 2001). There are 3 principles concluded as following: (1) the original data cannot be destroyed; (2) the evidence must be consistent with the original data; (3) original evidence cannot be damaged during analysis process.

The hard disk, one kind of very important electronic evidence, is the most common storage medium in the computer use. There are two key research issues in the Computer forensics evidence preservation technology, one is Effective protection of the original hard disk from the writ-
ing of new data or the destruction of evidence in the process of evidence collection, the other is maintenance of hard disk data integrity. The write-protect devices, which is used to avoid the original hard disk written by the new data, can solve the problem of evidence authenticity may be damaged. However, in terms of data integrity protection, although many domestic and foreign researchers gradually have proposed some hard drive integrity verification program, and increasing the efficiency of verification perfect in view of the traditional mechanical hard disk, the data integrity protection issues have not been resolved according to the increasingly widespread use of SSD (Solid State Disk), by far.

Common data integrity verification method is to use cryptographic hash algorithm (Hash) produce data summaries for comparison. Nowadays in the lots of commercial computer forensics software (Chow, et al., 2005; GS), the common calibration procedures for hard disk integrity check are as follows. First, with cryptography Hash algorithm (Hash), the hard disk data arranged in a certain order to calculate and obtain a hash value H. Then have the value stored for integrity checking. In order to ensure that the hard drive data is not affected by various factors in the mining process and then changed, the hard disk data will be calculated once again to get another forensics hash value h. If before and after hash values result in the same, representative of hard disk data did not change, whereas that of hard disk data has changed, the digital evidence extracted from the hard disk is also thought to be unreliable. Obviously, for this calibration method, if a Bit of hard drive data changes, forensic results will be regarded as invalid. Computer forensics is a time-consuming and lengthy process, including the influence of various factors. So, if the hard disk of individual sector data change in the process, not only unable to isolate this part of the data but also continue to use other unchanged data, it will lead to the whole Forensics process invalid (Adelstien, et al., 2006).

There is another more intuitive integrity check algorithm which utilizes the hash calculation, too. The difference between it and previous one is that the hashing algorithm is applied to every sector of a hard disk instead of the entire hard disk (Jiang, et al., 2008). This algorithm has the advantage that calibration results can accurately pick out which part is the real one to affect the integrity of the sector. However, the calibration data, which the algorithm needs to store, stored in the hard disk is significantly increased with the increase of hard disk storage.

For a capacity of 250GB hard drive for instance, assuming that the hash value stored length of 128 bits (for example: MD5), you will take about 7GB of space to store the generated hash algorithm calibration data. In addition, for the remote computer forensics or computer forensics in cloud computing environment, transporting such a large number of calibration values will add extra burden to the system.

Harbour (2002) proposed Piecewise Hash. The Piecewise Hash calculate more than one checksum for each file, so that a single error can just affect a checksum, and at the same time it reduces the effect of single error on the integrity of the entire image file.

The piecewise hash algorithm was improved by Kornblum (2006) who proposed a hash algorithm based on CTPH, short for Context Triggered Piecewise Hash, which can identify different modified versions of the same file. The above algorithm is just for small volume file integrity checking so that it focuses on functionality while ignoring the efficiency issue. With Piecewise Hash the time required for the calculation of calibration data is increased with which the amount of data n is $O(n\log n)$ growth. If the algorithm is used for the treatment of hard disk data integrity problems, the time required to calculate calibration data will quickly grow as the increase of the hard drive capacity to the scale which is not suitable for practical use.

Combinatorial Group Testing abbreviated to CGT is a method proposed by Fang (2011) and other parallel-group for checking hard disk integrity efficiently. Combinatorial Group Testing can be widely used to deal with error correction of large quantities of data. But the precondition is that wrong number of samples is far less than
Scheduling Algorithms and Radio Resource Management in WiMAX

RFID Technology in Business and Valuation Methods
[www.igi-global.com/chapter/rfid-technology-in-business-and-valuation-methods/115147?camid=4v1a](www.igi-global.com/chapter/rfid-technology-in-business-and-valuation-methods/115147?camid=4v1a)

A Semantic Approach to Designing Information Services for Smart Museums
[www.igi-global.com/article/a-semantic-approach-to-designing-information-services-for-smart-museums/180292?camid=4v1a](www.igi-global.com/article/a-semantic-approach-to-designing-information-services-for-smart-museums/180292?camid=4v1a)