基于有限射影几何的细粒度数据完整性检验方法

陈 龙[1],娄晓会[1],王国胤[2]
(1.重庆邮电大学计算机取证研究所,重庆 400065; 2.重庆邮电大学计算机科学与技术研究所,重庆 400065)

摘 要: 细粒度的数据完整性检验方法在实现完整性检验的同时可以对少数的错误对象进行准确和高效的隔离,从而避免因偶然错误或个别篡改造成整体数据失效的灾难性后果.对细粒度数据完整性检验问题进行了总结和分类,给出了总体的研究思路.为了提高细粒度数据完整性检验方法的错误指示效率,基于有限射影几何原理构造了一种新的数据完整性指示码.其思想是将有限射影几何空间中点与线的关联关系映射为 Hash 与数据对象之间的监督关系,实现 Hash 之间完全的均匀交叉.分析了码的主要性能.分析和实验结果表明该码可以准确指示多个错误并且具有更高的压缩率.

关键词: 计算机取证; Hash; 数据完整性; 有限射影几何
中图分类号: TP309 **文献标识码:** A **文章编号:** 0372-2112 (2011) 12-2850-06

An Integrity Check Method for Fine-Grained Data Based on Finite Projective Geometry

CHEN Long[1], LOU Xiao-hui[1], WANG Guo-yin[2]
(1. *Institute of Computer Forensics, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*;
2. *Institute of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*)

Abstract: Fine-grained data integrity checking methods could isolate a portion of corrupted data segments and assure the integrity of other data at the same time. They could mitigate the disastrous effect that all the data become invalid caused by accidental errors or intentional forging modification. The issues of integrity checking for fine-grained data are summarized at first. Based on the finite projective geometry theory, a novel data integrity checking method (integrity indication code) is proposed to improve the error indication efficiency. The basic idea of the method is mapping the relationship of points and lines in finite projective geometry to the check relationship of hashes and data objects, and achieving uniform crossover of all the hashes. The performances analysis and experimental results show that this code is effective with higher compression ratio.

Key words: computer forensics; hash; data integrity; finite projective geometry

1 引言

计算机取证工作的难点之一是证明取证人员所收集到的证据没有被修改过.人们常采用 Hash 检验、数字签名、时间戳等多种技术固定证据[1,2],采用预先存储、获取方式来解决证据缺乏的问题[2,3].Hash 检验在计算机取证领域的重要应用之一是计算并存储取证映像(完全复制件)的 Hash 值,从而保证分析用的副本的完整性.取证映像的数据量往往很大,而海量数据处理是计算机取证面临的另一难题[1].由于 Hash 函数的特性,任何数据变化都会影响到全体数据的可用性和可信性,所以对海量数据(比如取证映像)的完整性保护不能只停留在整体是否可靠、未被修改的层面上,理想的解决方法是隔离已经发生变动的数据而继续使用其它未改变的数据[4].细粒度的数据完整性检验成为支持改善海量数据可用性的一种重要手段,文献[5]在开发取证映像工具时基于直观的磁头、柱面、扇区区分法,从不同的角度对磁盘数据进行交叉校验,实现了一种基本的细粒度完整性检验方案,增强了磁盘数据的完整性与可用性.

细粒度数据完整性检验的一般性问题可归纳为:如何高效地判断每个细粒度数据对象是否具有完整性?细粒度是一个相对于传统所关注的数据对象大小的概

收稿日期:2010-03-20;修回日期:2011-07-06
基金项目:国家自然科学基金(No.60573068);重庆市教委科学技术研究项目(No.KJ110505);重庆邮电大学博士启动基金(No.A2009-25)

念,例如目录中的单个文件、一个文件的独立分片——小数据块、磁盘的扇区级物理存储块[6]或数据流中的数据包.这样一来,伴随着海量数据处理本身的问题,完整性检验 Hash 数据也成为大规模数据,而这些数据无法使用压缩技术进行压缩,这将给存储和网络传输带来极大不便[7].

借鉴纠错编码思想[8]可以通过交叉完整性检验,在低出错率条件下使用少量 Hash 监督大量数据对象,实现细粒度完整性检验的同时减少 Hash 数据量[7].相对于每个数据对象使用 1 个 Hash 监督的方案,这种交叉检验方案体现为 Hash“压缩”.陈龙等在讨论细粒度数据完整性指示编码基本性质的基础上已分别构造了针对单错、多错的细粒度数据完整性指示编码[7,9,10].其中,文献[10]提出的复数旋转指示码可以准确指示多个错误,但在需要准确指示的错误数稍多时,压缩效果较差.计算机取证的实际案例越来越复杂,在需要处理细粒度的较大规模数据甚至海量数据时,现有的多错完整性指示码的压缩率较低,不能满足实际需要.

本文基于有限射影几何原理,利用其中的点与线的关联关系构造了一种新的细粒度多错完整性指示码,实现 Hash 之间的完全均匀交叉,从而达到更高的压缩率.

2　完整性指示码

2.1　完整性指示码的基本概念

例 1,设 $X_1, X_2, X_3, X_4, X_5, X_6$ 表示 6 个数据对象,采用 4 个 Hash 监督这 6 个数据对象,使用如下的监督关系(交叉检验):

$$\begin{cases} X_1 \parallel X_2 \parallel X_4 = h_1 \\ X_1 \parallel X_3 \parallel X_5 = h_2 \\ X_2 \parallel X_3 \parallel X_6 = h_3 \\ X_4 \parallel X_5 \parallel X_6 = h_4 \end{cases} \tag{1}$$

式(1)中的“$\parallel$”表示将数据对象连接成一个数据流,“=”表示将左端的数据流进行单向 Hash 运算,等式右端 h_1, h_2, h_3, h_4 表示 Hash.该监督方案可准确指示一个错误.在需要进行完整性检验时采用相同顺序处理数据对象,按式(1)重新生成 Hash,与事先存储的 Hash 进行比较以判断数据对象是否变化.例如,h_1, h_2 与其原值不相符,而 h_3, h_4 无变化则明确表明其他数据对象没有出错,具有完整性,而 X_1 出错.

定义 1(监督矩阵)　令 $N=\{1,2,\cdots,n\}$ 为 n 个需要进行完整性监督的数据对象的编号集合,$M=\{M_1, M_2,\cdots,M_m\}$ 为 m 个 Hash 所监督的数据对象编号集构成的集合.监督矩阵 $\boldsymbol{A}$ 是按如下方式定义的 $m\times n$ 的 0、1 矩阵.

$$A=(a_{ij}),1\leqslant i\leqslant m,1\leqslant j\leqslant n \tag{2}$$

其中:　$a_{ij}=\begin{cases}1, & \text{若 } j\in M_i\\ 0, & \text{若 } j\notin M_i\end{cases}$.

监督矩阵 $\boldsymbol{A}$ 表达了 Hash 与其监督对象之间的监督关系,$a_{ij}=1$ 的含义是第 j 个数据对象受第 i 个 Hash 监督.

定义 2(完整性指示码)　设有一种 m 个 Hash 监督 n 个数据对象的监督方案,利用所生成的 m 个 Hash 进行完整性检验,如果在检验时能准确指示任意 t 个出错对象,而在 $n\geqslant t+1$ 时至少存在 $t+1$ 个错误数据对象的组合无法准确指示,其中受监督次数最多的某个数据对象受到 k 个 Hash 监督 ($k\geqslant 1$),那么把该监督方案称为一个完整性指示码,记为$[n,m,t,k]$.具体地设计一种监督方案就是设计一个编码.码的压缩率 η 为数据对象数 n 和使用的 Hash 个数 m 之比.

定义 3(错误放大率)　利用完整性指示码 $C=[n,m,t,k]$进行完整性检验时,若实际出现的错误数 x 大于编码设计时可准确指示的错误数 t,则可能出现将正常对象判定为出错对象的情况,即指示出的出错对象数大于 x,这种现象称错误放大.由于错误对象的分布不同,实际指示错误数也可能不同,考察 x 个错误对象的所有分布,可得其平均数.指示错误对象的平均数与实际出错数 x 的比值称为错误放大率,记为 $\beta(x)$.由于码 C 能准确指示 t 个错误,所以出现 $t+1$ 个错误时的 $\beta(x)$最能体现码的主要错误放大特性,$x=t+1$ 时的 $\beta(x)$简记为 β,称为码 C 的基准错误放大率.特别地,规定 $\beta(0)=1$.

2.2　完整性指示编码需要研究的问题

数据毁坏的问题需要使用数据安全的机制解决,本文不讨论.对于少量数据对象出错的情况,按完整性出错原因可以分为偶然错和人为错,人为错又可分为单方人为错与多方人为错(合谋欺骗).单方人为错指数据的实际控制者一方自行根据自己的目的修改数据.多方人为错指相关数据的实际控制者根据共同的目的各自对相应数据进行修改以体现一致性.单方人为错在不考虑数据毁坏等因素的情况下将主要是少量的数据删改(包括 Hash)或伪造,此时单方人为错与偶然错特性相似.根据问题的复杂程度以及现有研究工作解决问题思路,可将数据对象出错分为单错和多错两类.按照完整性检验方案中 Hash 监督的数据对象是否交叉及压缩效果可将完整性指示编码方案分为无交叉检验、可压缩的交叉检验和无压缩的交叉检验三类.组合这些需求,完整性指示编码需要研究的问题可归纳为表 1.

表 1 中划线部分表明不必要或不存在该方案.单方

人为错对数据完整性的影响与偶然错类似,可用相同方法解决.现有研究结果已构造了组合单错完整性指示码、超方体单错完整性指示码、复数旋转指示码等三种方案[7,9,10],分别适于解决其中的三个问题.其它问题尚待研究.本文针对偶然错条件下的均匀交叉 Hash 方案对应的问题,构造新的编码,给出解决方案.

表 1 细粒度数据完整性检验问题

完整性检验方案		偶然错		人为错	
		单错	多错	单方人为错	多方人为错
无交叉检验	单 Hash 检验	传统方案		同偶然错方案	——
	多重 Hash 检验	——			多备份 Hash
交叉检验且 $m<n$	Hash 均匀交叉	组合单错指示码	本文方案		待研究
	Hash 独立分组	超方体单错指示码	复数旋转指示码		待研究
	其他方式	待研究			待研究
交叉检验且 $m\geqslant n$		——			待研究

说明:m 为 Hash 个数,n 为数据对象个数

3 基于有限射影几何的完整性指示码

3.1 有限射影几何及其点线关系

设 GF(q)是含有 q 个元素的有限域,其中 q 为素数或素数幂.一个 $d+1$ 维向量$(\xi_0,\xi_1,\cdots,\xi_d)$叫做一个点,如果一切$(\rho\xi_0,\rho\xi_1,\cdots,\rho\xi_d)$都与$(\xi_0,\xi_1,\cdots,\xi_d)$表示同一个点,其中 $\rho\neq 0,\rho\in \mathrm{GF}(q),\xi_i\in \mathrm{GF}(q),i=0,1,\cdots,d$,$d$ 为正整数.我们把全部这样的点的集合叫做有限域 GF(q)上的 d 维射影几何,记做 PG(d,q)[8,11].有限射影几何空间 PG(d,q)中可定义不同维度的子空间,直线是其中的 1 维子空间[8,11].由 ρ 的取值可知每个点对应 $q-1$ 个向量,所以有限射影几何 PG(d,q)中点的总数为:

$$m=\frac{q^{d+1}-1}{q-1}=\sum_{i=0}^{d}q^i \tag{3}$$

若 $P_1(\xi_0,\xi_1,\cdots,\xi_d)$和 $P_2(\zeta_0,\zeta_1,\cdots,\zeta_d)$是 d 维射影几何的两个点,令$(\rho_0\xi_0+\rho_1\zeta_0,\rho_0\xi_1+\rho_1\zeta_1,\cdots,\rho_0\xi_d+\rho_1\zeta_d)$为 P_1、P_2 连线上的点,其中 $\rho_0,\rho_1\in \mathrm{GF}(q)$,且不全为零.由于 ρ_0,ρ_1 不能同时为零,故 ρ_0,ρ_1 共有 q^2-1 种组合方式,又由于每个点对应 $q-1$ 个向量,所以每条直线上点的数目为:

$$k=\frac{q^2-1}{q-1}=q+1 \tag{4}$$

考虑从 PG(d,q)中选取两个线性无关的向量的选法个数.第一个向量可以是 PG(d,q)中的任一非零向量,有 $q^{d+1}-1$ 种选法;第二个向量需与第一个向量线性无关,有 $q^{d+1}-q$ 种选法.因此,从 PG(d,q)中选取两个线性无关的向量的选法个数为:$(q^{d+1}-1)(q^{d+1}-q)$.

类似地,对于每条直线,从中选取两个线性无关的向量的方法共有$(q^2-1)(q^2-q)$种.

所以,PG(d,q)中不同直线的条数为:

$$n=\frac{(q^{d+1}-1)(q^{d+1}-q)}{(q^2-1)(q^2-q)} \tag{5}$$

现在考虑 PG(d,q)中包含某一固定点 $P_1(\xi_0,\xi_1,\cdots,\xi_d)$的直线的条数:这样的直线中的两个点中的一个可以由 P_1 充当,另外一个点在 P_1 以外选取,有 $q^{d+1}-q$ 种选法.

类似地,对每一条这样的直线,有 q^2-q 种选取 P_1 点以外的一个符合要求的点的方法.

因此,PG(d,q)中包含某一固定点的直线的条数为:

$$r=\frac{q^{d+1}-q}{q^2-q}=\sum_{i=0}^{d-1}q^i \tag{6}$$

由于 PG(d,q)中任意两条线最多交于一点,因此任意一对相异点只属于一条直线.

3.2 PG(d,q)中直线的生成

GF(q)上的 $d+1$ 维向量空间中的向量与GF(q^{d+1})中的元素存在着一一对应关系,所以 PG(d,q)的点也可用 GF(q^{d+1})的非零元表示.设 α 为 GF(q^{d+1})的本原元,则 PG(d,q)的点可以用 α 的前 m 个方次表示,即 $\alpha^0,\alpha^1,\cdots,\alpha^{m-1}$.于是给定两个点 $\alpha^{c_0},\alpha^{c_1}$,可以由下式得到由这两个点确定的直线.

$$\alpha^w=\rho_0\alpha^{c_0}+\rho_1\alpha^{c_1} \tag{7}$$

其中 $\rho_0,\rho_1\in \mathrm{GF}(q)$,且不全为零.当 ρ_0,ρ_1 取不同的值时,可以得到由这两个点确定的直线上的其余 $q-1$ 个点.因为点 α^w 与指数 w(整数)一一对应,为方便起见,可使用指数表示该点,直线则使用$\{c_0,c_1,\cdots,c_q\}$的形式表示.由此得到的直线称为一条初始直线.由于 PG(d,q)上的直线可分为若干组,每组直线可由一条初始直线用简化的方法循环生成.即从一条初始直线开始,通过加整数 1,2,…,$m-1$ 到该直线上的各个点(对 m 取模的整数加法运算),即可得到其它与该直线相循环的直线.例如,某空间的一条初始直线为$\{0,6,8\}$,则下一条直线为$\{1,7,9\}$.

初始直线构造算法如下:

步骤 1 初始化整数集合 $S=\varnothing$;

步骤 2 通过式(7)获得有限射影几何空间 PG(d,q)中的一条初始直线$\{c_0,c_1,\cdots,c_q\}$,为简化运算,设定 $c_0=0$,而 c_1 为不属于 S 且满足 $c_1\leqslant\frac{m}{2}$的自然数,如果不存在这样的 c_1,则已形成所有的初始直线;

步骤 3 根据直线$\{c_0,c_1,\cdots,c_q\}$,由下式计算相应的差值集合 s;

$$s=\{(c_i-c_j+m)\bmod m \mid i\neq j, i,j=0,1,\cdots,q\} \quad (8)$$

步骤 4　将差值集合 s 并入集合 S, $S=S\cup s$;

步骤 5　重复步骤 2 到 4,直至形成所有的初始直线.

3.3　基于有限射影几何的完整性指示码

3.3.1　基于有限射影几何的完整性指示码的构造

定义 4(有限射影完整性指示码)　设有 $n=\frac{(q^{d+1}-1)(q^{d+1}-q)}{(q^2-1)(q^2-q)}$ 个(q 是素数或素数幂, $d\geqslant 3$)待检验的数据对象,需要准确指示 t 个错误.将 m 个 Hash 依次对应于 PG(d, q)上的 $m=\sum_{i=0}^{d}q^i$ 个点, n 个数据对象依次对应于 PG(d, q)上的各条直线,由点线关系来确定 Hash 监督关系.由于每条直线包含 $q+1$ 个点,则每条直线对应 $q+1$ 个 Hash,它们共同监督该直线对应的数据对象.该监督方案共有 m 个 Hash,并使 n 个数据对象与 m 个 Hash 形成均匀交叉:每个数据对象同时被 $q+1$ 个 Hash 监督,每个 Hash 监督 $r=\frac{q^{d+1}-q}{q^2-q}$ 个数据对象,形成均匀交叉的有限射影完整性指示码

$$C=[n,m,t,k]=\left[\frac{(q^{d+1}-1)(q^{d+1}-q)}{(q^2-1)(q^2-q)},\sum_{i=0}^{d}q^i,q,q+1\right]$$

定理　有限射影完整性指示码 $C=[n,m,t,k]=\left[\frac{(q^{d+1}-1)(q^{d+1}-q)}{(q^2-1)(q^2-q)},\sum_{i=0}^{d}q^i,q,q+1\right]$ 可准确指示任意 q 个错误.

证明　在有限射影完整性指示码中,每个数据对象对应一条直线,每个 Hash 对应一个点,每条直线上有 $k=q+1$ 个点.设 q 个出错数据对象对应的直线分别为 L_i, $i=0,1,\cdots,q-1$,相应出错 Hash 对应的点集合为 M.

(1)对于每一个出错数据对象,它所对应直线上的点必然在集合 M 中,所以 q 个出错对象都能被指示出.

(2)对任一未出错的数据对象,设其对应的直线为 L,由于每两条直线的交点至多为 1,所以 L 与 L_i, $i=0,1,\cdots,q-1$ 这 q 条直线的交点至多为 q,即 L 上的点至多有 q 个在集合 M 中,则 L 上至少有一点不在 M 中,即有该数据对象参与计算的 Hash 至少有一个未出错,所以,该数据对象不会被判定为出错,即不会出现误检.

综上,有限射影完整性指示码 $C=[n,m,t,k]=\left[\frac{(q^{d+1}-1)(q^{d+1}-q)}{(q^2-1)(q^2-q)},\sum_{i=0}^{d}q^i,q,q+1\right]$ 可准确指示任意 q 个错误,证毕.

3.3.2　Hash 生成

设数据对象个数为 n,错误指示能力为 t.选取适当的 d、q(q 为素数或素数幂),构造射影几何 PG(d, q),使得 $n\leqslant\frac{(q^{d+1}-1)(q^{d+1}-q)}{(q^2-1)(q^2-q)}$.首先生成一组初始直线,然后循环生成所有直线,并将其依次编号为 0,1,…, $n-1$.数据对象与 Hash 的对应:把数据对象依次编号为 0,1,…, $n-1$,将数据对象与直线按照编号依次对应,每个数据对象由对应直线上的点所对应的 Hash 共同监督.各个数据对象与 Hash 形成均匀交叉,每个数据对象同时受 k 个 Hash 监督,每个 Hash 监督 r 个数据对象.

采用并发计算方式,依次读入各个数据对象,同时推进所有 Hash 的计算过程,每个数据对象只需读入一次.具体步骤如下:

步骤 1　根据需要,选择合适的 d, q 值;

步骤 2　构造一组初始直线;

步骤 3　初始化所有 Hash, $j=0$;

步骤 4　读入第 j 个数据对象;

步骤 5　根据初始直线生成第 j 条直线,

步骤 6　将第 j 个数据对象与第 j 条直线对应,确定需要参与计算的 k 个 Hash;

步骤 7　分别取出各个 Hash 对应的中间结果或者初值计算所有 k 个 Hash,将中间结果再存入相应的 Hash 中;

步骤 8　$j=j+1$,重复步骤 4 到 7 直到数据对象处理完;

步骤 9　输出最后计算出的一组 Hash 值.

3.3.3　Hash 检验

如果一个数据对象出错,则有该数据对象参与计算的 Hash 必然全部出错,否则,如果有一个 Hash 无错,则该数据对象无错.

具体检验步骤如下:

步骤 1　选取与原 Hash 数据生成时相同的参数,采用相同的 Hash 计算方式生成 Hash 矩阵;

步骤 2　依次比较各个 Hash 值,统计不相符的 Hash 数量,并记录其编号;

步骤 3　如果不相符的 Hash 值个数少于 k 个,则没有数据对象出错,所有数据都具有完整性,结束.否则继续下一步;

步骤 4　依次从记录的不相符的 Hash 中选取两个,如果由这两个 Hash 对应的点所生成的直线已经检查过,则重新从不相符的 Hash 中选取两个;否则,继续下一步;

步骤 5　根据式(8)得到由这两个 Hash 对应的点所确定的直线.检查该直线上其余点,看其对应的 Hash 值是否出错,如果全部出错,则该直线对应的数据对象出错.

步骤 6　重复步骤 4、5,验证其它数据对象是否出错.

3.3.4　性能分析

3.3.4.1　压缩率分析

(1)有限射影完整性指示码的压缩率

有限射影完整性指示码的压缩率为：

$$\eta = \frac{\sum_{i=0}^{d-1} q^i}{q+1} \tag{9}$$

显然，对某一固定维数 d，q 越大时可以准确指示的错误数越多，检验的数据对象越多，压缩率也越高. 为了直观地看出压缩率变化，取不同的维数 3,4,5,6，有限射影完整性指示码的压缩率变化如图 1 所示.

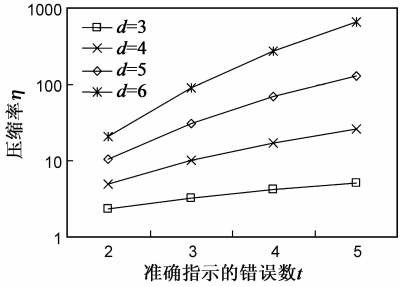


图1 有限射影完整性指示码的压缩率变化

由图 1 可以看出，随着 d 和 q 的增长，码的压缩率增长很快，整体而言，码的压缩率很高. 当指错能力 t 一定时，d 值越大，Hash 个数 m 和待检对象数量 n 越大，有限射影完整性指示码的压缩率越大. 当 d 值一定时，t 值越大，Hash 个数 m 和待检对象数量 n 越大，码的压缩率也越大. 整体而言，有限射影完整性指示码的压缩率较高，可以轻易实现百倍乃至千倍的压缩.

(2)有限射影完整性指示码与复数旋转指示码的压缩率比较

复数旋转指示码[p^2, $p(t+1)$, t, $t+1$]的压缩率[10]为(p 为素数或素数幂，t 为准确指示的错误数)：

$$\eta' = \frac{p}{t+1} \tag{10}$$

对某一固定错误数 t 和相同的 Hash 个数 m，有限射影完整性指示码和复数旋转指示码的压缩率之比为：

$$\Delta\eta_1 = \frac{\eta}{\eta'} = \frac{\sum_{i=0}^{d-1} t^i}{p} = \frac{(t+1)\sum_{i=0}^{d-1} t^i}{\sum_{j=0}^{d} t^j} = (t+1) - \frac{t+1}{\sum_{i=0}^{d} \frac{1}{t^i}} \tag{11}$$

由于 $d \geqslant 3$、$t \geqslant 2$，所以 $\Delta\eta_1 > 1$.

对某一固定错误数 t 和相同的待检对象数量 n，有限射影完整性指示码和复数旋转指示码的压缩率之比为：

$$\Delta\eta_2 = \frac{\eta}{\eta'} = \frac{\sum_{i=0}^{d-1} t^i}{p} = \frac{\sum_{i=0}^{d-1} t^i}{\sqrt{\frac{(t^{d+1}-1)(t^{d+1}-t)}{(t^2-1)(t^2-t)}}} = \sqrt{\Delta\eta_1} \tag{12}$$

则 $\Delta\eta_2 > 1$.

显然，两种条件下有限射影几何码与复数旋转指示码的压缩比都大于 1. 当 t 一定时，d 值越大，Hash 个数 m 和待检对象数量 n 越大，两种完整性指示码的压缩比也越大.

3.3.4.2 错误放大率分析

采用程序抽样的方法估计错误放大率 β.

设出错数为 $q+1$，在 $n = \frac{(q^{d+1}-1)(q^{d+1}-q)}{(q^2-1)(q^2-q)}$ 个数据中选 $q+1$ 个出错数有 $C(n, q+1)$种组合. 当 n 较小时，对 $C(n, q+1)$种组合进行枚举来确定 β；当 n 较大时，用抽样的方式确定 β，实验设定的抽样次数为 50 万次.

对不同的指错能力 $t=2,3,4,5,7$，不同的维数 $d=3,4,5$，基准错误放大率 β 的变化趋势如图 2 所示.

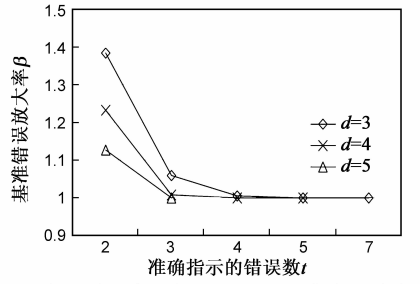


图2 有限射影完整性指示码的基准错误放大率

由图 2 可以看出，有限射影完整性指示码的基准错误放大率 β 较低，当指错能力 t 一定时，β 随 d 值的增大而减小并且趋近于 1，当维数 d 一定时，β 随 q 值增大而减小并且趋近于 1. 当实际出错数为 $e=t+1$ 时，有限射影完整性指示码在绝大多数情况下都能正确指示出这 $t+1$ 个错误.

可见，有限射影完整性指示码在需要准确指示的错误数较少或维数较低时，检验的数据对象较少，错误放大率稍大，而在需要准确指示的错误数较多或维数较高时，检验的数据对象较多，错误放大率较低并且趋近于 1.

4 结论

本文对细粒度数据完整性检验问题进行了总结和分类，分析了需要进一步研究的问题. 基于有限射影几何原理，提出了一种新的细粒度数据完整性指示码，解决了上述分类问题中的一个. 该码利用有限射影空间中的点与直线的关系进行编码设计，在保证设定的错误指示能力条件下，实现 Hash 之间完全的均匀交叉. 理论分析和实验结果表明，该码能够准确指示多个错误，与复数旋转指示码相比，该码具有更高的压缩率，在数据量较大并且对错误指示能力要求较高的场合能够很好地发挥作用，可以大大节省 Hash 数据的存储空间和

传输 Hash 数据需要的网络带宽,适用于取证数据量比较大的场合.

参考文献

[1] Golden G. Richard III, Vassil Roussev. Next-generation digital forensics[J]. Communications of the ACM, 2006, 49(2): 76 – 80.

[2] 孙波,孙玉芳,等.电子数据证据收集系统保护机制的研究与实现[J].电子学报,2004,32(8):1374 – 1380.
Sun Bo, Sun Yufang, et al. Research and implementation of the protection mechanism for digital evidence collecting system [J]. Acta Electronica Sinica, 2004, 32(8): 1374 – 1380. (in Chinese)

[3] 王文奇,苗凤君,等.网络取证完整性技术研究[J].电子学报,2010,38(11):2529 – 2534.
Wang Wenqi, Miao Fengjun, et al. The research on integrity technique of network-based forensic[J]. Acta Electronica Sinica, 2010, 38(11): 2529 – 2534. (in Chinese)

[4] The Common Digital Evidence Storage Format Working Group. Standardizing digital evidence storage[J]. Communication of the ACM, 2006, 49(2): 67 – 68. .

[5] Zoe L Jiang, Lucas C K Hui, S M Yiu. Improving disk sector integrity using K-dimension Hashing[A]. Advances in Digital Forensics IV[C]. Kyoto, Japan: Springer, 2008. 87 – 98.

[6] Roussev V, Chen Y, et al. Md5bloom: forensic filesystem hashing revisited[J]. Digital Investigation, 2006, 3(sl): 82 – 90.

[7] 陈龙,王国胤.一种细粒度数据完整性检验方法[J].软件学报,2009,20(4):902 – 909.
Chen Long, Wang Guoyin. An integrity check method for fine-grained data[J]. Journal of Software, 2009, 20(4): 902 – 909. (in Chinese)

[8] 靳蕃,陈志.组合编码原理及应用[M].上海:上海科学技术出版社,1995.

[9] 陈龙,方新蕾,王国胤.系列单错完整性指示码及其性能分析[J].计算机科学,2009,36(6):97 – 100.
Chen Long, Fang Xinlei, Wang Guoyin. One error integrity indication codes and performance analysis[J]. Computer Science, 2009, 36(6): 97 – 100. (in Chinese)

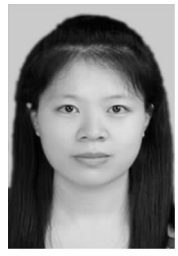
[10] 陈龙,方新蕾,王国胤.基于复数旋转码的细粒度数据完整性指示方法[J].西南交通大学学报,2009,44(5):667 – 671.
Chen Long, Fang Xinlei, Wang Guoyin. Integrity check method for fine-grained data based on complex rotary codes [J]. Journal of Southwest Jiaotong University, 2009, 44(5): 667 – 671. (in Chinese)

[11] 朱勇,张宝富,李玉权.光正交码的有限射影几何设计方法[J].通信学报,1999,20(1):28 – 33.
Zhu Yong, Zhang Baofu, Li Yuquan. A method of construction optical orthogonal codes from finite projective geometries[J]. Journal on Communications, 1999, 20(1): 28 – 33. (in Chinese)

作者简介

陈　龙　男,教授,中国电子学会高级会员、中国计算机学会高级会员.1970 年生于重庆.1992 年、2000 年和 2009 年分别在华东师范大学、电子科技大学和西南交通大学获得工学学士、工学硕士和工学博士学位.主要研究领域为计算机取证,网络安全,云计算安全,智能信息处理.

E-mail: chenlong@cqupt.edu.cn



娄晓会　女,1985 年生,硕士研究生,研究方向为信息安全

E-mail: lxh04130106@yahoo.cn