

整数的最佳带符号二进制表示的随机生成算法

李忠^{1,2} 张永华¹

(宜宾学院计算机科学与技术研究所 宜宾 644000)¹ (宜宾学院计算机与信息工程学院 宜宾 644000)²

摘要 整数的带符号数字(BSD)表示广泛应用于计算机算术、密码学、数字信号处理等领域。给定整数有多种最佳带符号二进制表示。对整数的最佳带符号二进制表示的性质进行研究,给出了整数的最佳带符号二进制表示随机生成算法,所得算法能快速产生给定整数的随机最佳带符号二进制表示。

关键词 整数,带符号二进制表示,最佳带符号二进制表示,随机生成算法

中图法分类号 TP301.6 **文献标识码** A

Random Generation Algorithm of Optimal Binary Signed Digit Representation of Integer

LI Zhong^{1,2} ZHANG Yong-hua¹

(Institute of Computer Science & Technology, Yibin University, Yibin 644000, China)¹

(School of Computer & Information Engineering, Yibin University, Yibin 644000, China)²

Abstract Binary Signed Digit(BSD) representation of an integer is widely used in computer arithmetic, cryptography and digital signal processing. A given integer can have several optimal BSD representations. This paper studied the properties of the optimal BSD representation of an integer, and presented an generation algorithm, which can rapidly generate a random optimal BSD representation of a given integer.

Keywords Integer, Binary signed digit representation, Optimal binary signed digit representation, Random generation algorithm

对于正整数 k ,称 $k=k_{n-1}2^{n-1}+\dots+k_12^1+k_02^0$ (其中 $k_i\in\{-1,0,1\},0\leq i<n$)为 k 的长度为 n 的带符号二进制(BSD; Binary Signed Digit)表示。整数的BSD表示在计算机算术、密码学、数字信号处理等领域得到广泛应用。一个长度为 n 比特的整数有多种BSD表示。文献[1]对整数的BSD表示数进行了研究,给出计算整数的给定长度的BSD表示数的递归算法。文献[2]进一步对整数的BSD表示数的性质进行了研究,给出了计算整数的给定长度的BSD表示数的非递归算法,提升了计算效率。文献[1]还对整数的所有BSD表示的生成及随机BSD表示的生成问题进行了研究。在给定整数的BSD表示中,汉明重量最小的表示称为整数的最佳BSD表示。文献[3]讨论了给定整数的最佳BSD表示的长度及计数问题。文献[4]首次对给定整数的所有最佳BSD表示的生成进行了研究。文献[5,6]进一步讨论了给定整数的所有最佳BSD表示的生成问题。本文进一步研究整数的最佳BSD表示的性质,提出给定整数的随机最佳BSD表示算法,所得算法能快速产生给定整数的随机最佳BSD表示。

1 整数的最佳BSD表示

到目前为止,已提出多种整数的最佳BSD表示算法^[7-9],其中,非相邻形式(NAF; Non-Adjacent Form)表示^[7]是最典型的最佳BSD表示,它是椭圆曲线密码体制(ECC; Elliptic

Curve Cryptosystem)无预计算标量乘法运算的首选表示算法^[10]。整数的NAF表示算法如算法1所示。

算法1 计算正整数的NAF表示

输入:正整数 $k=(k_{i-1}\dots k_1k_0)_2$

输出:NAF(k)

1. $c_0=0, k_{i+1}=0, k_i=0$;
2. for i from 0 to l do
 - 2.1 $c_{i+1}=\lfloor(c_i+k_i+k_{i+1})/2\rfloor$;
 - 2.2 $u_i=c_i+k_i-2c_{i+1}$;
3. return($u_lu_{l-1}\dots u_1u_0$)

对于整数的NAF表示有如下性质:

定理1^[7] 设 k 是正整数,则

- (1) k 有唯一的NAF表示,简记为NAF(k);
- (2)NAF(k)没有两个连续的数字是非零的,且是 k 在数字集 $D_s=\{-1,0,1\}$ 上所有BSD表示中具有最小汉明重量的表示;
- (3)NAF(k)的长度至多比 k 的基本二进制表示的长度大1;
- (4)数字集 $D_s=\{-1,0,1\}$ 上所有长度为 l 的NAF表示中,平均汉明密度约为1/3。

定义1 设给定整数 k 的NAF表示为 $NAF(k)=(u_{r-1}\dots u_2u_1u_0)_2, u_{r-1}\neq 0$,则

(1)NAF(k)中最长的连续比特串 $\{01\}^+$ 及 $\{0i\}^+$ 称为一

本文受四川省教育厅重点科研项目(13ZA0196),宜宾学院博士科研启动金项目(2012B16)资助。

李忠(1963—),男,博士,副教授,CCF会员,主要研究方向为密码学、信息安全, E-mail: lz806859@163.com; 张永华(1967—),女,高级实验师,主要研究方向为计算机应用技术。

个块;

(2) NAF(k) 中至少两个连续的块组成的最长比特串称为一条链。

例如, 整数 $k=3017253757243$ 的 NAF 表示为:

$$k = (010i0i0000i01000001010i0i01010000101000i0i)_2 \quad (1)$$

它有 9 个块:

$$\begin{array}{cccccccc} \overline{01} & \overline{010} & \overline{1000} & \overline{01} & \overline{010000} & \overline{0101} & \overline{01010} & \overline{101} & \overline{0101000} & \overline{0101} & \overline{00} \\ B_1 & B_2 & B_3 & B_4 & B_5 & B_6 & B_7 & B_8 & B_9 & & \\ \overline{0101} & & & & & & & & & & \\ B_9 & & & & & & & & & & \end{array} \quad (2)$$

有 3 条链:

$$\begin{array}{ccccccc} \overline{0101010000} & \overline{01010000} & \overline{01010101010000101000} & \overline{10} \\ L_1 & L_2 & L_3 & \\ \overline{1} & & & \end{array} \quad (3)$$

其中, i 代表 -1 (以下相同)。我们有以下结论:

定理 2 设给定整数 k 的 NAF 表示为 $NAF(k) = (u_i \dots u_2 u_1 u_0)_2$, 则

- (1) 只有链能引起整数 k 的最佳 BSD 表示的变化;
- (2) 链 $0101 \dots 010iR$ 与链 $010iR$ 对于整数 k 的最佳 BSD 表示的影响相同;
- (3) 链 $0i0i \dots 0i01R$ 与链 $0i01R$ 对于整数 k 的最佳 BSD 表示的影响相同。

其中, R 为相应比特串 (以下类似)。

例如, 式 (1) 为整数 $k=3017253757243$ 的一种最佳 BSD 表示, 只有式 (3) 中的链 L_1-L_3 能引起 k 的最佳 BSD 表示的变化, 对于 k 的任意最佳 BSD 表示 $(u_{44} \dots u_2 u_1 u_0)_2$, 除与式 (3) 中 L_1-L_3 对应比特位可能不相同外, 其余部分完全一样。

定理 3 设 R 是链, 令 $U = (R)_2$, 有:

- (1) 若 $\dot{R}1$ 是 U 的最佳 BSD 表示, 则 $\dot{R}101$ 是 $\dot{U} = (R01)_2$ 的最佳 BSD 表示;
- (2) 若 $\dot{R}i$ 是 U 的最佳 BSD 表示, 则 $\dot{R}i01$ 、 $\dot{R}0i$ 是 $\dot{U} = (R01)_2$ 的最佳 BSD 表示;
- (3) 若 $\dot{R}1$ 是 U 的最佳 BSD 表示, 则 $\dot{R}10i$ 、 $\dot{R}011$ 是 $\dot{U} = (R0i)_2$ 的最佳 BSD 表示;
- (4) 若 $\dot{R}i$ 是 U 的最佳 BSD 表示, 则 $\dot{R}i0i$ 是 $\dot{U} = (R0i)_2$ 的最佳 BSD 表示。

定理 2、定理 3 的结论比较明显, 限于篇幅其证明从略。

定理 4 设 $0u_0 0u_2 \dots 0u_{2m}$ 是链, 令 $U = (0u_0 0u_2 \dots 0u_{2m})_2$, $U_i = (0u_0 0u_2 \dots 0u_{2i})_2$, $S_{i,1}$ 、 $S_{i,i}$ 分别表示 U_i 的以 '1'、'-1' 结尾的最佳 BSD 表示所构成的集合 ($0 \leq i \leq m$), 有:

- (1) 若 $u_{2i} = 1$, 则 $S_{i,1} = \{R01 | R \in S_{i-1,1}\} \cup \{R01 | R \in S_{i-1,i}\}$, $S_{i,i} = \{R00i | R0i \in S_{i-1,i}\}$;
- (2) 若 $u_{2i} = -1$, 则 $S_{i,1} = \{R0011 | R01 \in S_{i-1,1}\}$, $S_{i,i} = \{R0i | R(S_{i-1,1}) \cup \{R0i | R \in S_{i-1,i}\}\}$;
- (3) $S_{m,1} \cup S_{m,i}$ 构成 U 的最佳 BSD 表示。

证明: 由链的定义可知, $u_i \in \{-1, 1\}$, 且 U 及 U_i 的最佳 BSD 表示的最后一比特一定是 '1' 或 '-1' ($0 \leq i \leq m$), 显然有 (3) 的结论成立。

(1) 若 $u_{2i} = 1$, 对于任意 $\dot{R}1 \in S_{i-1,1}$ 、 $\dot{R}i \in S_{i-1,i}$, 由最佳 BSD 表示的定义及定理 3 有:

$\dot{R}101$ 、 $\dot{R}i01$ 、 $\dot{R}0i$ 是 U_i 的最佳 BSD 表示, 且只有这些表示才是 U_i 的最佳 BSD 表示, 所以, $S_{i,1} = \{R01 | R \in S_{i-1,1}\} \cup \{R01 | R \in S_{i-1,i}\}$, $S_{i,i} = \{R00i | R0i \in S_{i-1,i}\}$ 。

(2) 若 $u_{2i} = -1$, 对于任意 $\dot{R}1 \in S_{i-1,1}$ 、 $\dot{R}i \in S_{i-1,i}$, 由最佳 BSD 表示的定义及定理 3 有:

$\dot{R}10i$ 、 $\dot{R}011$ 、 $\dot{R}i0i$ 是 U_i 的最佳 BSD 表示, 且只有这些表示才是 U_i 的最佳 BSD 表示, 所以, $S_{i,1} = \{R0011 | R01 \in S_{i-1,1}\}$, $S_{i,i} = \{R0i | R \in S_{i-1,1}\} \cup \{R0i | R \in S_{i-1,i}\}$ 。

2 整数的最佳 BSD 表示的随机生成

一般情况下, 给定整数的最佳 BSD 表示不唯一, 用 $\lambda(k, n)$ 表示整数 $k \in [0, 2^n - 1]$ 的长度为 n 的不同 BSD 表示数, 用 $N(k)$ 表示 k 的最佳 BSD 表示数。对于 NIST 推荐的 ECC 基域 $GF(2^{192} - 2^{64} - 1)$ 的随机元素 $k = (9DA1893935071B91AD72F63E38FE488585D06025A612FF90)_{16} \in GF(2^{192} - 2^{64} - 1)$ 有 $\lambda(k, 193) = 106970910741389689572820341693667$, $N(k) = 27648$ 。

由定理 2 可知, 只有整数 k 的链能影响 k 的最佳 BSD 表示。由定理 4 可知, 在链 L 中, 可以通过反复使用: $10i$ 与 011 相互替换、 $i01$ 与 $0i$ 相互替换而得到新的最佳 BDS 表示。

由定理 1 可知, NAF(k) 是 k 的最佳 BSD 表示, 而且没有连续的两比特的值是非 '0' 的。根据定理 2、定理 4, 可以按以下方式获得给定整数 k 的随机最佳 BSD 表示:

(1) 利用算法 1 计算 NAF(k);

(2) 从左向右扫描 NAF(k), 当出现 $10i$ 或 $i01$ 时, 随机决定是否将其转换为 011 或 $0i$;

(3) 重复 (2) 的过程, 直到扫描完整个 NAF(k) 为止, 便得到 k 的随机最佳 BSD 表示。

具体算法如算法 2 所示。

算法 2 整数的随机最佳 BSD 表示的生成

输入: 整数 k

输出: k 的随机最佳 BSD 表示 $u = (u_1 \dots u_1 u_0)_2$

1. 利用算法 1 计算 $NAF(k) = (k_1 \dots k_1 k_0)_2$;
2. $k_{-1} = 0, k_{-2} = 0, i = 1$;
3. while $i > 0$ do
 - 3.1 while $i > 0$ and $k_i = 0$ do
 - 3.1.1 $u_i = k_i, i = i - 1$;
 - 3.2 if $k_i * k_{i-2} \neq -1$ then
 - 3.2.1 $u_i = k_i, u_{i-1} = k_{i-1}, i = i - 2$;
 - 3.3 else
 - 3.3.1 $x \in_R \{0, 1\}$;
 - 3.3.2 if $x = 0$ then
 - 3.3.2.1 if $k_i = 1$ and $k_{i-2} = -1$ then
 - 3.3.2.1.1 $u_i = 0, u_{i-1} = 1, k_{i-2} = 1, i = i - 2$;
 - 3.3.2.2 else
 - 3.3.2.2.1 $u_i = 0, u_{i-1} = -1, k_{i-2} = -1, i = i - 2$;
 - 3.3.3 else
 - 3.3.3.1 $u_i = k_i, u_{i-1} = k_{i-1}, i = i - 2$;
 - 3.4 if $i = 0$ then $u_0 = k_0$;
5. return $u = (u_1 \dots u_1 u_0)_2$

对于 $k=3017253757243$, 利用算法 2 可得其随机最佳 BSD 表示 $(0010110000i01000001010i0i0i010000101000i0i)_2$ 。

(下转第 293 页)

- [20] Kaufman L, Rousseeuw P. Finding Groups in Data: An Introduction to Cluster Analysis[M]. Wiley, 1990
- [21] Duda R O, Hart P E, Stork D G. Pattern Classification, Second Edition[M]. A Wiley-Interscience Publication, 2001
- [22] Peng Jing, et al. A new similarity computing method based on concept similarity in Chinese text processing[J]. Science in China Series F: Information Sciences, 2008, 51(9): 1215-1230
- [23] Wang Hai-xun, Wang Wei, Yang Jiong, et al. Clustering by Pattern Similarity in Large Data Sets[C]//Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data. 2002; 394-405
- [24] Tibshirani R, et al. Estimating the number of clusters in a data set via the gap statistic[J]. J. R. Statist. Soc. B, 63, part 2, 2001; 411-423
- [25] Fraley C, Raftery A E. How Many Clusters? Which Clustering Method? Answer Via Model-Based Cluster Analysis[J]. The Computer Journal, 1998, 41(8): 578-588
- [26] Sun Hao-jun, Wang Sheng-rui, Jiang Qing-shan. FCM-Based Model Selection Algorithms for Determining the Number of Clusters[J]. Pattern Recognition, 2004(37): 2027-2037
- [27] Halkidi M, Batistakis Y, Vazirgiannis M. On Clustering Validation Techniques[J]. Journal of Intelligent Information Systems, 2001, 17(2/3): 107-145
- [28] Fraley C, Raftery A E. Model-Based Clustering, Discriminant Analysis, and Density Estimation[J]. Journal of the American Statistical Association, 2002, 97(458): 611-631
- [29] LEroux B G. Consistent Estimation of a Mixing Distribution[J]. The Annals of Statistics, 1992, 20(3): 1350-1360
- [30] Baya A E, Granitto P M. ISOMAP based metrics for clustering[J]. Inteligencia Artificial, 2008, 12(37): 15-23
- [31] Tenenbaum J B, de Silva V, Langford J C. A Global Geometric Framework for Nonlinear Dimensionality Reduction[J]. Science, 2000, 290: 2319-2323
- [32] Hyvarinen A, Oja E. Independent Component Analysis: algorithms and applications[J]. Neural Networks, 2000, 13: 411-430
- [33] Alter O, Brown P O, Botstein D. Singular value decomposition for genome-wide expression data processing and modeling[J]. PNAS, 2000, 97(18): 10101-10106
- [34] Kim P M, Tidor B. Subsystem identification through dimensionality reduction of large-scale gene expression data[J]. Genome Res, 2003, 13(7): 1706-1718
- [35] Murtagh F, Starck J L, Berry M W. Overcoming the Curse of Dimensionality by Means of the Wavelet Transform[J]. The Computer Journal, 2000, 43: 107-120
- [36] De Winter J, Wagemans J. Segmentation of object outlines into parts: A large-scale integrative study[J]. Cognition, 2006, 99: 275-325
- [37] Freixenet J, Munoz X, Raba D, et al. Yet Another Survey on Image Segmentation: Region and Boundary Information Integration[C]//ECCV2002. LNS2352, 2002; 408-422
- [38] 李瑞, 邱玉辉. 基于离散点的蚁群聚类算法的研究[J]. 计算机科学, 2005, 32(6): 111-113
- [39] 田铮, 李小斌, 句彦伟. 谱聚类的扰动分析[J]. 中国科学 E 辑: 信息科学, 2007, 37(4): 527-543
- [40] Barbara B, Chen Ping. Using the Fractal Dimension to Cluster Datasets[C]//Proc. of the 6th ACM SIGKDD Int'l Conf. on Knowledge discovery and data mining (KDD-2000). ACM Press, 2000; 260-264
- [41] 杨博, 刘大有, Liu Ji-ming, 等. 复杂网络聚类方法[J]. 软件学报, 2009, 20(1): 54-66
- [42] 张莉, 周伟达, 焦李成. 核聚类算法[J]. 计算机学报, 2002, 25(6): 587-590
- [43] Al-Shammary D, Khalil I, Tari Z. A distributed aggregation and fast fractal clustering approach for SOAP traffic[J]. Journal of Network and Computer Applications, 2014, 41: 1-14
- [44] Basu B, Srinivas V V. Regional flood frequency analysis using kernel-based fuzzy clustering approach[J]. Water Resources Research, 2014, 50(4): 3295-3316
- [45] Li Xiang, Wong Hau-san, Wu Si. A fuzzy minimax clustering model and its applications[J]. Information Sciences, 2012, 186(1): 114-125

(上接第 283 页)

结束语 本文对给定整数的最佳 BSD 表示的性质进行了深入研究, 给出了计算给定整数的随机最佳 BSD 表示的算法, 所得算法能快速产生给定整数的随机最佳 BSD 表示。后续工作中, 将利用本文的结论设计抗边信道攻击的 ECC 快速标量乘法算法。

参 考 文 献

- [1] Ebeid N, Hasan M A. On binary signed digit representations of integers[C]//Design Code Cryptogr. 2007, 42: 43-65
- [2] 李忠, 彭代渊. 整数的带符号二进制表示数的快速计算[J]. 计算机应用, 2012, 32(11): 3121-3124
- [3] Wu T, Zhang M, Du H, et al. On optimal binary signed digit representations of integers[J]. Applied Mathematics, 2010, 25(3): 331-340
- [4] Ganesan P, Manku G S. Optimal routing in Chord[C]//Proc. 15th ACM-SIAM Symposium on Discrete Algorithms (SODA 2004). 2004; 169-178
- [5] Sawada J. A Gray code for binary subtraction[C]//2nd Brazilian Symposium on Graphs, Algorithms and Combinatorics (GRACO 2005). 2005
- [6] Manku G S, Sawada J. A Loopless Gray Code for Minimal Signed-Binary Representations[C]//Brodal G S, Leonardi S, eds. ESA 2005. LNCS 3669, 2005; 438-447
- [7] Hankerson D, Menezes A, Vanstone S. Guide to elliptic curve cryptography[M]. Springer-Verlag Professional Computing Series, 2004
- [8] Avanzi R M. A note on the signed sliding window integer recoding and a left-to-right analogue[C]//Handschuh H, Hasan A, eds. SAC 2004. LNCS 3357, 2004; 130-143
- [9] Joye M, Yen S M. Optimal left-to-right binary signed-digit recoding[J]. IEEE Transactions on Computers, 2000, 49: 740-748
- [10] Okeya K, Schmidt-samoak C, Spahn, et al. Signed binary representations revisited[C]//Advances in Cryptology-CRYPTO'04. LNCS 3152, 2004; 123-139