Nova geração da representação D-ária de um número racional

Valdemar C. da Rocha Jr. e Marcos Müller Vasconcelos

***Resumo*— Este artigo apresenta uma nova maneira de obtenção da representação D-ária para qualquer número racional positivo menor que 1, em termos de séries geométricas. Em contraste com a representação D-ária convencional, onde um termo futuro somente pode ser calculado após todos os termos anteriores serem conhecidos, a nova técnica fornece o conhecimento de qualquer termo futuro independentemente do conhecimento de termos anteriores. Como resultado, o armazenamento de qualquer representação D-ária pode ser feita mais eficientemente armazenando sempre um número finito de parâmetros. Este procedimento encontra aplicação em criptografia, em teoria da informação e na geração de números aleatórios.**

***Palavras-Chave*— Criptografia, números racionais, expansões em série, codificação homofônica.**

***Abstract*— This paper presents a new way of obtaining the D-ary representation of any positive rational number less than 1, in terms of geometrical series. Contrasting with the conventional D-ary representation, where a future term can only be computed after all its predecessors are known, the new technique provides knowledge about any future term independent of knowledge of previous terms. As a result the storage of any D-ary representation can be done more efficiently by storing a finite number of parameters. This procedure finds application in cryptography, information theory and generation of random numbers.**

***Keywords*— Cryptography, rational numbers, series expansions, homophonic coding.**

I. Introdução

Sistemas criptográficos de chave-secreta, nos quais a mensagem cifrada contém pouca ou nenhuma redundância, são mais difíceis de serem quebrados [1]. Substituição homofônica [1]-[3] é uma técnica criptográfica usada para reduzir a redundância de uma mensagem a ser cifrada, ao custo de uma expansão do texto-claro. Na substituição homofônica, cada letra da mensagem original é representada por um substituto ou *homofonema* em um alfabeto maior, para formar a mensagem em texto-claro que será então cifrada. Seja D um número inteiro positivo. Em uma substituição homofônica D-ária a representação de cada letra por palavras código D-árias é alcançada pela expansão em base D da probabilidade da letra correspondente, isto é, através da expansão de um número racional positivo λ, menor que 1, e então representando cada homofonema por uma palavra código D-ária na decomposição D-ária de λ. Uma necessidade similar de realizar expansões D-árias também ocorre na geração de números aleatórios [4]-[6]. Freqüentemente a expansão de uma probabilidade numa dada base D possui um número infinito de termos, ou seja, o símbolo da fonte associado a esta probabilidade possui um número infinito de palavras código para representar os seus homofonemas. Pensava-se então que esta seria uma grande desvantagem para o uso prático da substituição homofônica, pois a mesma iria requerer um dicionário com um número infinito de palavras código. A representacção D-ária aqui apresentada permite que, para probabilidades que são números racionais, o dicionário infinito de homofonemas possa ser construído a partir um dicionário finito de palavras auxiliares.

Exemplo 1: Seja $D = 2$ e considere uma fonte binária $S = \{a, b\}$ onde as probabilidades de ocorrência dos símbolos a e b são, respectivamente, $P_S(a) = 1/4$ e $P_S(b) = 3/4$. A codificação homofônica binária de S produz um homofonema v_1 para representar a letra a e dois homofonemas, v_2 e v_3, para representar b, onde $P(v_1) = 1/4,\ P(v_2) = 1/2$ e $P(v_3) = 1/4$. As palavras-código $0, 10$ e 11 podem ser usadas para representar v_1, v_2 and v_3, respectivamente.

O objetivo deste trabalho é introduzir uma nova e eficiente maneira de obter a representação D-ária de um número racional positivo menor que 1, no sentido de que qualquer termo específico da representação possa ser gerado sem o conhecimento obrigatório dos termos anteriores. Deve-se observar que a extensão destes resultados para cobrir a repesentação de qualquer número racional é trivial. Isto segue pois qualquer número maior que 1 sempre pode ser decomposto como uma soma de um número inteiro positivo mais um número racional positivo menor que 1. A decomposição de números inteiros positivos como uma soma de potências de $D,\ D \geq 2$, é imediata, assim como também lidar com o caso de números negativos.

Na seção 2 é apresentado um procedimento para obter de forma eficiente a representação D-ária de um número inteiro positivo menor do que 1, assim como também são apresentados alguns exemplos. Este artigo é encerrado na seção 3, com alguns comentários.

II. Representação D-ária

A expressão "representação D-ária de um número racional positivo menor do que 1" significa a representação de tal número como uma soma de potências negativas de um número inteiro positivo D, onde cada termo pode ter multiplicidade no máximo $D - 1$. Em geral, a forma convencional de se calcular um termo futuro em uma representação D-ária requer o conhecimento de todos os termos anteriormente calculados. Esta exigência pode impor limitações severas em muitas aplicações.

Proposição 1: Qualquer número racional menor do que 1 pode ser expandido em base D como uma soma de um número

Valdemar C. da Rocha Jr. e Marcos Müller Vasconcelos¸ Grupo de Pesquisa em Comunicações - CODEC, Departmento de Eletrônica e Sistemas, Universidade Federal de Pernambuco. E-mails: vcr@ufpe.br, mmv@ee.ufpe.br. Este trabalho foi parcialmente financiado pelo CNPq, Projeto 305226/2003-7 e PIBIC-30161.

finito de termos distintos (não-recursivos) mais um número infinito de termos (recursivos) pertencentes a um número finito de séries geométricas infinitas, cujos primeiros termos são potências negativas de D de multiplicidade no máximo $D-1$, e as razões são potências negativas de D.

Demonstração: Seja m/n um número racional, onde $m < n$, m e n são números inteiros positivos e primos entre si, ou seja, o máximo divisor comum entre m e n é igual 1 e será denotado por $\text{mdc}(m,n) = 1$.

1) Suponha que a decomposição de n em fatores primos contém apenas fatores (repetidos ou não) que aparecem na decomposição de D em fatores primos. Segue que n divide D^r, sendo r o maior expoente entre os fatores primos da decomposição de n. Portanto

$$\frac{m}{n} = \frac{m}{n}\frac{D^r}{D^r} = \frac{h}{D^r} = \frac{h_0}{D^r} + \frac{h_1}{D^{r-1}} + \ldots + \frac{h_{r-1}}{D},$$

na qual $h = mD^r/n$ é um número inteiro positivo menor que D^r e $h_0, h_1, \ldots, h_{r-1}$ são os coeficientes da representação de h na base D. Conclui-se que m/n neste caso expande na base D como uma soma de no máximo r termos não-recursivos distintos.

2) Suponha agora que a decomposição de n em fatores primos contém fatores relativamente primos a D, cujo produto é denotado por n_1, além de fatores (repetidos ou não) que aparecem na decomposição de D em fatores primos, cujo produto é denotado por n_2. Segue que n divide D^{r+s}, sendo r o maior expoente entre os fatores primos da decomposição de n_2 e sendo $\lambda(n_1) = s$ a função de Carmichael [7, págs. 275-277], ou seja, s é o menor número inteiro positivo para o qual $D^s - 1$ é divisível por n_1. Segue que

$$\frac{m}{n} = \left[\frac{m}{n_1}\frac{(D^s-1)}{(D^s-1)}\right]\left[\frac{1}{n_2}\frac{D^r}{D^r}\right] = \frac{h_1}{(D^s-1)}\frac{h_2}{D^r}. \quad (1)$$

$$\frac{h_1/D^s}{1-D^{-s}} = \left(\frac{h_{10}}{D^s} + \frac{h_{11}}{D^{s-1}} + \ldots + \frac{h_{1,s-1}}{D}\right)\sum_{j=0}^{\infty} D^{(-s)j}$$

$$\frac{h_2}{D^r} = \frac{h_{20}}{D^r} + \frac{h_{21}}{D^{r-1}} + \ldots + \frac{h_{2,r-1}}{D},$$

na qual $h_1 = m(D^s-1)/n_1$ e $h_2 = D^r/n_2$ são números inteiros positivos menores que D^s e D^r respectivamente, e $h_{10}, h_{11}, \ldots, h_{1,s-1}$ e $h_{20}, h_{21}, \ldots, h_{2,r-1}$ são os respectivos coeficientes da representação de h_1 e de h_2 na base D. Segue de (1) que m/n expande na base D como uma soma de no máximo $r+s$ séries geométricas infinitas tendo D^{-s} como razão comum. No caso de $h_1 > 1$ procedemos a separação de h_1 em dois termos, sendo um deles um número inteiro positivo h_1' e o outro uma fração própria h_1''/n_1, conforme ilustrado no Exemplo 2.

■

Exemplo 2: Seja $D = 2$. A representação binária de $1/5$ e $7/10$ de acordo com a Proposição 1 é como segue.

a) Pela proposição 1 a representação binária de $1/5$ produz $s = 4$ e $h_1 = m(2^s-1)/n = 3$, assim segue que

$$\begin{aligned} 1/5 &= \frac{(3/16)}{1-2^{-4}} \\ &= \frac{(1/8)}{1-2^{-4}} + \frac{(1/16)}{1-2^{-4}} \\ &= (1/8)\sum_{j=0}^{\infty}(1/16)^j + (1/16)\sum_{j=0}^{\infty}(1/16)^j. \end{aligned}$$

b) Seja $m/n = 7/10$. Pela Proposição 1 a representação binária de $7/10$ produz $n_1 = 5,\ n_2 = 2, r = 1$ e $s = 4$, assim segue que

$$\begin{aligned} 7/10 &= \left(\frac{1}{2}\right)\left(\frac{7}{5}\right) \\ &= \left(\frac{1}{2}\right)\left[1+\frac{2}{5}\right] \\ &= \frac{1}{2} + \frac{1}{5} \\ &= \frac{1}{2} + (1/8)\sum_{j=0}^{\infty}(1/16)^j + (1/16)\sum_{j=0}^{\infty}(1/16)^j, \end{aligned}$$

onde na última igualdade foi usada a representação de $1/5$ do Exemplo 2a).

Exemplo 3: Seja $D = 3$. A representação ternária de $7/9,\ 5/8$ e $7/15$ de acordo com a Proposição 1 é como segue.

$$\begin{aligned} \frac{7}{9} &= \frac{2.3+1}{3^2} = 2.3^{-1} + 3^{-2} \\ \frac{5}{8} &= \frac{\left[(3^2-1)/3^2\right]5}{\left[(3^2-1)/3^2\right]8} \\ &= \frac{5/9}{1-3^{-2}} \\ &= \frac{(3+2)/9}{1-1/9} \\ &= \frac{1/3}{1-1/9} + \frac{2/9}{1-1/9} \\ &= (1/3)\sum_{i=0}^{\infty}(1/9)^i + (2/9)\sum_{i=0}^{\infty}(1/9)^i, \\ \frac{7}{15} &= (1/3)(7/5) \\ &= (1/3)(1+2/5) \\ &= 1/3 + (1/3)(2/5) \\ &= 1/3 + \frac{(2/5)(80/81)}{80/81} \\ &= 1/3 + \frac{32/81}{1-1/81} \\ &= 1/3 + \frac{27/81 + 3/81 + 2/81}{1-80/81} \\ &= 1/3 + (1/3)\sum_{i=0}^{\infty}(1/81)^i + \\ &\quad (1/27)\sum_{i=0}^{\infty}(1/81)^i + (2/81)\sum_{i=0}^{\infty}(1/81)^i. \end{aligned}$$

III. APLICAÇÃO

O procedimento descrito pode ser aplicado em diversas situações de interesse prático. Uma das aplicações consiste na geração de números aleatórios D-ários com distribuição de probabilidade uniforme, a partir de uma fonte S discreta sem memória, com símbolos cujas probabilidades são números racionais [8]. A técnica consiste em decompor na base D a probabilidade de cada símbolo de S e de representar cada termo desta decomposição por uma palavra D-ária de um código unicamente decodificável. As seqüências geradas por este esquema são compostas por variáveis aleatórias D-árias independentes e identicamente distribuídas com distribuição uniforme.

Exemplo 4: Seja $D = 6$ e consideremos uma fonte binária $S = \{a, b\}$ onde as probabilidades de ocorrência dos símbolos a e b são, respectivamente, $P_S(a) = 7/20$ e $P_S(b) = 13/20$. Usando o método proposto, obtemos a expansão das probabilidades na base 6 conforme indicado a seguir.

$$\begin{aligned}
\frac{7}{20} &= \frac{7}{2^2 \cdot 5} \cdot \frac{6^2}{6^2} \cdot \frac{6-1}{6-1} \\
&= \frac{7 \cdot 3^2}{6^2} \cdot \left(\frac{1/6}{1-1/6} \right) \\
&= \left(1 + 4 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6^2} \right) \cdot \frac{1}{6} \sum_{i=0}^{\infty} \left(\frac{1}{6} \right)^i \\
&= \frac{1}{6} \sum_{i=0}^{\infty} \left(\frac{1}{6} \right)^i + 4 \cdot \frac{1}{6^2} \sum_{i=0}^{\infty} \left(\frac{1}{6} \right)^i + 3 \cdot \frac{1}{6^3} \sum_{i=0}^{\infty} \left(\frac{1}{6} \right)^i
\end{aligned}$$

$$\begin{aligned}
\frac{13}{20} &= \frac{13}{2^2 \cdot 5} \cdot \frac{6^2}{6^2} \cdot \frac{6-1}{6-1} \\
&= \frac{13 \cdot 3^2}{6^2} \cdot \left(\frac{1/6}{1-1/6} \right) \\
&= \left(3 + \frac{1}{6} + 3 \cdot \frac{1}{6^2} \right) \cdot \frac{1}{6} \sum_{i=0}^{\infty} \left(\frac{1}{6} \right)^i \\
&= 3 \cdot \frac{1}{6} \sum_{i=0}^{\infty} \left(\frac{1}{6} \right)^i + \frac{1}{6^2} \sum_{i=0}^{\infty} \left(\frac{1}{6} \right)^i + 3 \cdot \frac{1}{6^3} \sum_{i=0}^{\infty} \left(\frac{1}{6} \right)^i
\end{aligned}$$

Um possível código 6-ário empregado para representar os termos da decomposição terá cardinalidade infinita e consiste das palavras indicadas na Tabela I.

TABELA I
CODIFICAÇÃO D-ÁRIA DE FONTE BINÁRIA.

Símbolo	Probabilidade	Palavra
a	$1/6$	0
	$1/36$	40
	$1/36$	41
	$1/36$	42
	$1/36$	43
	$1/36$	44
	$1/108$	530
	$1/108$	531
	$1/108$	532
	$1/108$	533
	$1/108$	534
	$1/108$	535
	$1/108$	540
	$1/108$	541
	$\vdots$	$\vdots$
b	$1/6$	1
	$1/6$	2
	$1/6$	3
	$1/36$	45
	$1/36$	50
	$1/36$	51
	$1/36$	52
	$1/108$	542
	$1/108$	543
	$1/108$	544
	$1/108$	545
	$1/108$	550
	$1/108$	551
	$1/108$	552
	$\vdots$	$\vdots$

IV. CONCLUSÕES

Este artigo se encerra com duas observações. Primeiramente, segue da proposição 1 que a representação D-ária $(D \geq 2)$ de um número racional positivo é completamente caracterizada por um número finito de parâmetros. Em segundo lugar, qualquer termo desta representação D-ária pode ser gerado daqueles parâmetros sem a necessidade da geração todos os termo anteriores. Os resultados apresentados neste artigo tem relevância em aplicações práticas em substituição homofônica [1]- [3] e na geração de números aleatórios [4] - [6].

REFERÊNCIAS

[1] H. N. Jendal, Y. J. B. Kuhn and J. L. Massey, "An information-theoretic approach to homophonic substitution", pp. 382-394 in *Advances in Cryptology-Eurocrypt'89* (Eds. J.-J. Quisquater and J. Vandewalle), Lecture Notes in Computer Science, No.434. Heidelberg and New York: Springer, 1990.

[2] V.C. da Rocha Jr. and C. Pimentel, "Binary-constrained homophonic coding", *International Symposium on Communication Theory and Applications*, Ambleside, U.K, 15-20 July, 2001, pp.263-268.

[3] M. Hoshi and T.S. Han, "Interval algorithm for homophonic coding", *IEEE Trans. Inform. Theory*, vol. IT-47, pp.1021-1031, March 2001.

[4] D.W. Knuth and A.C. Yao, "The complexity of random number generation", In J.F. Traub, editor,*Algorithms and Complexity: Recent Results and New Directions. Proceedings of the Symposium on New Directions and Recent Results in Algorithms and Complexity, Carnegie Mellon University, 1976*. Academic Press, New York, 1976.

[5] Julia Abrahams, "Generation of Discrete Distributions form Biased Coins", *IEEE Trans. Inform. Theory*, vol. IT-42, pp.1541-1546, September 1996.

[6] M. Hoshi and T.S. Han, "Interval algorithm for random number generation", *IEEE Trans. Inform. Theory*, vol. IT-43, pp. 599 -611, March 1997.

[7] H. Riesem, *Prime Numbers and Computer Methods for Factorization*, Birkhäuser, Basel, 1985.

[8] V.C. da Rocha Jr. and C. Pimentel, "On the generation of discrete distributions from a biased coin", *International Symposium on Information Theory ISIT*, Yokohama, Japan, 29 June - 4 July, 2003, p. 339.