Codificação Homofônica Universal Utilizando Codificação Diferencial e Entrelaçamento

Daniel da Rocha Simões, Jaime Portugheis e Valdemar C. da Rocha Jr.

***Resumo*— Um novo esquema para codificação homofônica universal é introduzido o qual combina propriedades da codificação diferencial e do entrelaçamento de símbolos. Como consequência, tanto a geração como a recuperação do texto-claro produzidos pelo esquema proposto são de fácil implementação. Testes para validação deste esquema foram realizados usando a suite de testes estatísticos NIST IR 6483 do *National Institute of Standards and Technology* e os respectivos resultados das simulações em computador são apresentados. Estes testes demonstram que o desempenho do esquema proposto é superior aos demais com os quais foi comparado, face à suite de testes do NIST e à expansão do texto-claro.**

***Palavras-Chave*— Criptografia, codificação de fonte, codificação homofônica universal.**

***Abstract*— A new homophonic coding scheme is introduced which combines properties of differential encoding and symbol interleaving. As a consequence, both the generation and the recovery of the clear-text produced by the proposed scheme are easy to implement. Validation tests for this scheme were performed using the statistical test suite NIST IR 6483 from the *National Institute of Standards and Technology* and the corresponding computer simulation results are presented. "These tests demonstrate that the performance of the proposed scheme is superior to that of those schemes to which it was compared, regarding the NIST test suite and the clear text expansion."**

***Keywords*— Cryptography, homophonic coding, universal source coding.**

I. Introdução

A criptografia é a área de estudos que cuida basicamente de problemas de integridade, autenticidade e de sigilo de dados, como fichas médicas de pacientes, resultados de experimentos genéticos, cadastros bancários, etc. Na prática, os dados a serem protegidos por meio da criptografia possuem, em geral, um comportamento estatístico muito diferente daquele apresentado por dados produzidos por uma fonte aleatória, ou seja, por uma fonte que produz símbolos estatisticamente independentes e uniformemente distribuídos (IID). Tal comportamento representa uma vulnerabilidade que, caso não tratada adequadamente, certamente poderá ser explorada por terceiros.

A codificação homofônica é uma técnica utilizada na criptografia para combater ataques que exploram desvios na estatística de uma fonte de dados (texto-claro), que idealmente deveria ser uma fonte IID. No passado, a excessiva redundância do texto cifrado resultou na quebra de alguns sistemas de cifragem, motivando assim o interesse atual pela codificação homofônica. A codificação homofônica consiste na substituição de cada símbolo da mensagem original por um ou mais símbolos, denominados homofonemas (palavra de origem grega, significando "do mesmo som"), pertencentes a um alfabeto maior, de forma a produzir símbolos estatisticamente independentes e uniformemente distribuídos, reduzindo assim a redundância na mensagem original. Na sua forma clássica, esse procedimento necessita do conhecimento prévio da estatística do texto-claro para realizar a codificação. Apesar desta técnica ser conhecida há muitos anos, foi apenas em 1988 que Günther [2], em um trabalho pioneiro, descreveu um algoritmo para a realização de codificação homofônica, no qual as palavras representando homofonemas podem ter comprimento variável.

Na maioria das aplicações práticas, em geral, não se tem *a priori* o conhecimento da estatística da fonte, de modo que procedimentos de codificação homofônica para fontes específicas tornam-se bastante ineficientes nesta situação. Massey sugeriu um sistema de codificação homofônica universal baseado em um esquema universal de codificação de fonte, ou seja, um esquema que não necessita do conhecimento *a priori* da estatística da fonte [3]. O esquema proposto por Massey foi analisado em [4] e testes para validação do mesmo foram realizados usando a suite de testes estatísticos NIST IR 6483. Através destes testes algumas imperfeições foram detectadas.

Neste trabalho é examinada a realização da codificação homofônica universal utilizando a codificação diferencial dos símbolos da fonte combinados com símbolos de uma fonte aleatória auxiliar e com um entrelaçador. Esta técnica tem o potencial de alcançar o mesmo nível de desempenho do esquema proposto em [4] com uma menor expansão do texto-claro.

Este artigo está organizado como descrito a seguir. Na Seção II, são revistas algumas noções básicas de criptografia e a motivação para o uso da codificação homofônica. A Seção III descreve o esquema proposto por Massey [3] e a Seção IV descreve o esquema introduzido em [4], discutindo vantagens e limitações dos mesmos. Na Seção V apresentamos uma nova proposta para a codificação homofônica universal, baseada no uso da codificação diferencial e entrelaçador. Na Seção VI descrevemos alguns detalhes dos testes estatísticos empregados e apresentamos resultados da simulação do esquema proposto. Na Seção VII concluímos este artigo com alguns comentários e observações sobre as vantagens e limitações do esquema de codificação universal proposto.

Daniel da Rocha Simões e Valdemar C. da Rocha Jr., Departamento de Eletrônica e Sistemas, Centro de Tecnologia e Geociências, Universidade Federal de Pernambuco, Recife, Brasil, Caixa Postal 7800, 50711-970, E-mail: drsdaniel82@gmail.com, vcr@ufpe.br. Jaime Portugheis, Departamento de Comunicações, UNICAMP, Campinas, Brasil, Caixa Postal 6101, 13083-970, E-mail:jaime@decom.fee.unicamp.br

II. MOTIVAÇÃO PARA O USO DA CODIFICAÇÃO HOMOFÔNICA

Cifras de chave-secreta não-expansivas são aquelas nas quais, para uma dada seqüência de números inteiros positivos $n_1, n_2, n_3, \cdots$, os primeiros n_i símbolos de texto-cifrado $Y_1, Y_2, \cdots, Y_{n_i}$ juntamente com a chave-secreta determinam de modo único os primeiros n_i símbolos do texto-claro correspondente $X_1, X_2, \cdots, X_{n_i}$, para $i = 1, 2, 3, \cdots$ [5]. Cifras sequenciais aditivas fornecem um exemplo de cifras de chave-secreta não expansivas, para as quais $Y_i = X_i \oplus Z'_i$, em que $Z'_1, Z'_2, Z'_3, \cdots$ denota a chave de sessão derivada da chave-secreta Z. Um outro exemplo é representado pelas cifras de bloco, de chave-secreta, para as quais tanto os blocos de texto-claro como os blocos de texto-cifrado têm o mesmo comprimento N. No caso das cifras sequenciais aditivas ocorre $n_i = i$ e no caso das cifras de bloco ocorre $n_i = iN$. As cifras de chave-secreta não-expansivas possuem a seguinte importante propriedade.

Proposição 2.1: Se uma seqüência de texto-claro X_n, cifrada por uma cifra de chave-secreta não-expansiva, for completamente aleatória então a seqüência de texto-cifrado Y^n é também completamente aleatória para qualquer escolha de z da chave Z. Além disso, Y^n é estatisticamente independente da chave-secreta Z.

Uma demonstração da Proposição 2.1 encontra-se na Referência [5]. Shannon definiu uma cifra *fortemente ideal* como aquela cifra para a qual $H(Z|Y^n) = H(Z)$, ou seja, para a qual a entropia da chave-secreta condicionada a n símbolos do texto-cifrado é igual à entropia da chave-secreta [6].

Corolário 2.1: Se uma seqüência de texto-claro X^n, cifrada por uma cifra de chave-secreta não-expansiva, for completamente aleatória, então o cripto-sistema é fortemente ideal, independentemente da distribuição de probabilidade da chave-secreta Z.

O Corolário 2.1 implica que um ataque de apenas texto-cifrado não consegue extrair informação alguma sobre a chave-secreta Z, não importando quantos símbolos do texto-cifrado sejam examinados. O objetivo da codificação homofônica é justamente o de transformar uma seqüência de símbolos de saída emitidos por uma fonte de informação não-aleatória em uma seqüência completamente aleatória, transformando qualquer cifra de chave-secreta não-expansiva em um cripto-sistema fortemente ideal.

III. ESQUEMA PROPOSTO POR MASSEY

Em 1988 C. G. Günther introduziu um esquema de substituição homofônica o qual transforma uma seqüência de texto-claro em uma seqüência de saída completamente aleatória, produzindo o que se denomina codificação homofônica perfeita [2]. Entretanto, uma limitação desta técnica é a necessidade do conhecimento *a priori* da estatística da fonte. A codificação homofônica universal, por outro lado, é de grande interesse na prática pelo fato de não necessitar de nenhum conhecimento *a priori* da estatística da fonte. Nosso ponto de partida na pesquisa sobre codificação homofônica universal foi o esquema proposto por Massey [3], o qual é ilustrado na Figura 1.

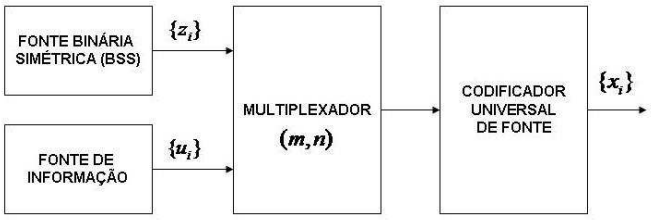

Fig. 1. Esquema de codificação homofônica universal proposto por Massey, no qual $\{z_i\}, \{u_i\}$ e $\{x_i\}$ denotam, respectivamente, as sequências de símbolos produzidos pela fonte binária simétrica, pela fonte de informação e pelo codificador universal de fonte.

Na Figura 1 a fonte binária de informação U emite símbolos u_i, $i = 1, 2, \cdots$, a fonte binária simétrica, denotada por BSS (do inglês *binary symmetric source*), denota um gerador aleatório que emite símbolos z_i. O multiplexador considerado é um dispositivo que produz sucessivamente na saída blocos contendo $m+n$ símbolos, sendo m símbolos da BSS, seguidos por n símbolos da fonte U. A saída do multiplexador é alimentada na entrada de um codificador universal de fonte apropriado.

Como a operação de multiplexação introduz não-estacionaridade, a seqüência de saída do multiplexador já não representa em geral uma fonte discreta estacionária e ergódica. No entanto, blocos de símbolos multiplexados cujo comprimento seja um múltiplo de $m + n$ produzem um processo cicloestacionário. A seqüência produzida na saída do codificador homofônico possui entropia dada por $H_\infty(V) = nH_\infty(U) + m$, na qual $H_\infty(U)$ denota a entropia da fonte de informação.

A sequência original $u_1, u_2, \ldots, u_i, \ldots$ produzida pela fonte de informação U é recuperada da seqüência $X = (x_1, x_2, \ldots, x_i, \ldots)$ produzida na saída do codificador homofônico universal, sem precisar conhecer o gerador aleatório empregado, processando X pelo respectivo decodificador universal de fonte e então descartando aqueles *bits* aleatórios que foram originados pela BSS.

O esquema de codificação universal proposto por Massey foi testado em [4] para verificar a aleatoriedade da sua sequência binária de saída, considerando fontes de informação distintas e várias combinações de valores para o par (m, n). Os resultados desses testes não foram satisfatórios, indicando a possibilidade de serem realizados melhoramentos. A Tabela III, mostrada no Apêndice A, ilustra resultados típicos obtidos nos testes. Essa foi a motivação para a proposta do esquema apresentado a seguir na Seção IV.

IV. ESQUEMA BASEADO NA CIFRA DE BLOCOS DESCARTÁVEIS E UM ENTRELAÇADOR

Ao multiplexar no tempo duas fontes de estatísticas distintas, o esquema ilustrado na Figura 1 deixa a cargo do codificador universal de fonte a dificílima tarefa de *aprender* uma estatística que muda o tempo todo, mantendo-se alternadamente em um valor durante m símbolos e num outro valor

durante n símbolos. Somando-se a essas dificuldades, temos o fato de que os m *bits* da fonte BSS são incompressíveis. Nos testes relatados em [4] essas dificuldades permaneceram, mesmo quando foram considerados codificadores universais com segmentos de comprimento fixo, operando com blocos de comprimento múltiplo de $m + n$ e uma entrada ciclo-estacionária.

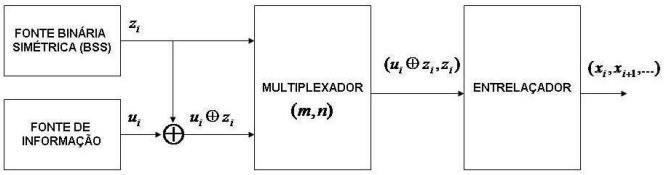

Fig. 2. Esquema para codificação homofônica universal baseado em cifra de blocos descartáveis e entrelaçador.

A fonte binária de informação U na Figura 2 emite símbolos u_i, $i = 1, 2, \cdots$, e é suposta ser estacionária e ergódica [3]. O bloco BSS denota uma fonte binária simétrica, isto é, um gerador aleatório com saída Z emitindo símbolos binários z_i que são igualmente prováveis e estatisticamente independentes (IID). O multiplexador considerado é um dispositivo o qual emite sucessivamente blocos de comprimento 2, contendo um símbolo z_i da BSS e um símbolo igual a $u_i \oplus z_i$. A saída do multiplexador é alimentada como entrada de um entrelaçador apropriado.

O esquema ilustrado na Figura 2, ao empregar a estrutura da cifra de blocos descartáveis (*one-time pad*), evita o problema de mudança de estatística na saída do multiplexador. Ou seja, conforme é explicado a seguir, como a seqüência $\{z_i\}$ é IID, a seqüência $\{u_i \oplus z_i\}$ também é IID.

Cifra de blocos descartáveis: Como é bastante conhecido, para $i = 1, 2, \ldots$, a cifra binária de blocos descartáveis (*one-time pad*) [6] tendo u_i como texto-claro e tendo z_i como chave-secreta, produz $u_i \oplus z_i$ como o correspondente texto-cifrado. A propriedade mais importante da cifra de blocos descartáveis, cuja prova se encontra em [6], é a seguinte.

Propriedade da cifra de blocos descartáveis: Se Z for completamente aleatória, isto é, se os valores assumidos pela variável aleatória Z obedecerem a uma distribuição de probabilidade uniforme e forem estatisticamente independentes, então a variável aleatória definida por $U \oplus Z$ será também completamente aleatória e não dependerá da distribuição de probabilidade de U.

Porém, como foi observado em [4],

$$\begin{aligned} P_{ZX}(z_i, u_i \oplus z_i) &= P_Z(z_i) P_{X|Z}(u_i \oplus z_i | z_i) \\ &= P_Z(z_i) P_U(u_i), \end{aligned} \quad (1)$$

que significa dizer que, em geral, Z e $U \oplus Z$ não satisfazem a condição exigida para independência estatística, isto é, a condição $P_{ZX}(z_i, u_i \oplus z_i) = P_Z(z_i) P_X(u_i \oplus z_i)$ em geral não é satisfeita, exceto para o caso não muito interessante no qual U já é completamente aleatório. Dessa forma, o multiplexador na Figura 2 emite pares $(u_i \oplus z_i, z_i)$ de símbolos binários em que ambos $u_i \oplus z_i$ e z_i são completamente aleatórios, entretanto não são necessariamente estatisticamente independentes. A fim de quebrar esta possível dependência estatística entre símbolos nos pares $(u_i \oplus z_i, z_i)$ foi introduzido um entrelaçador como ilustrado na Figura 2.

Na recepção, a seqüência original $u_1, u_2, \ldots$ de dígitos produzidos pela fonte de informação é facilmente recuperada a partir da seqüência recebida $x_1, x_2, \ldots$, usando apenas um desentrelaçador, um demultiplexador e um ou-exclusivo, sem a necessidade do conhecimento específico da fonte BSS. A Figura 3 ilustra um decodificador para o esquema de codificação universal proposto em [4].

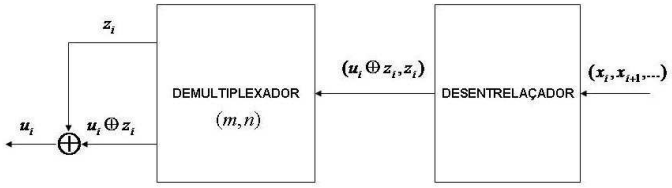

Fig. 3. Decodificador para o esquema de codificação universal ilustrado na Figura 2.

V. Esquema Baseado em Codificação Diferencial e Entrelaçador

O esquema apresentado na Seção IV produziu bons resultados quando submetido à suite de testes estatísticos do NIST. Sua implementação é simples e os resultados obtidos foram consideravelmente melhores que aqueles obtidos com o esquema proposto por Massey (vide Apêndice), empregando o codificador universal de fonte de Lynch-Davisson ou o codificador universal de fonte de Lempel-Ziv-Welch [4]. Na forma como foi apresentado, esse esquema produz na saída um arquivo cujo tamanho é o dobro do tamanho do arquivo original de texto-claro. Para contornar essa possível desvantagem, em [4] foi sugerido primeiramente comprimir a saída da fonte de informação U, antes de fazer a multiplexação.

Nesta seção apresentamos uma alternativa ao esquema introduzido em [4], a qual oferece como atrativo reduzir drasticamente a expansão do texto-claro original, sem sacrificar o desempenho do codificador, fato que é verificado por meio da suite de testes estatísticos do NIST. A Figura 4 ilustra o codificador homofônico universal proposto. A saída do codificador diferencial, denotada por $x_i, x_{i+1}, x_{i+2}, \ldots, x_{i+n}$, é formada por blocos de $n+1$ *bits* e em cada bloco o primeiro *bit* é oriundo da BSS. Um bloco de $n + 1$ *bits* é formado da seguinte maneira.

$$\begin{aligned} x_i &= z_i \\ x_{i+1} &= u_i \oplus x_i \\ x_{i+2} &= u_{i+1} \oplus x_{i+1} \\ &\vdots \\ x_{i+n} &= u_{i+n-1} \oplus x_{i+n-1}. \end{aligned}$$

Para um dado número de *bits* processados pelo entrelaçador, o valor de n foi determinado experimentalmente como sendo o maior número inteiro positivo para o qual o esquema é aprovado praticamente em todos os testes estatísticos e a partir do qual o esquema começa a falhar na grande maioria destes testes. Ou seja, foi observado um fenômeno de limiar no desempenho do esquema proposto, em função do valor de n.

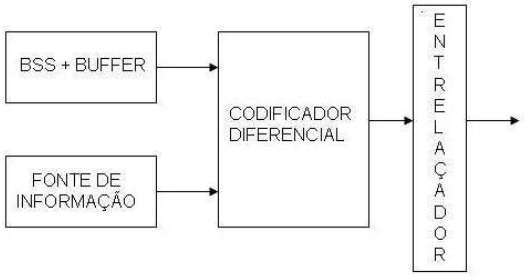

Fig. 4. Codificador para o esquema de codificação homofônica universal usando codificador diferencial e entrelaçador.

A i-ésima, $i = 1, 2, \ldots$, seqüência de n dígitos u_i, $u_{i+1}, \ldots, u_{i+n-1}$, contida em um bloco produzido pela fonte de informação, é facilmente recuperada a partir do bloco codificado correspondente, o qual após o desentrelaçamento é denotado por $x_i, x_{i+1}, \ldots, x_{i+n}$, operando da seguinte maneira,

$$u_{i+j} = x_{i+j+1} \oplus x_{i+j}, \quad \text{para} \quad j = 0, 1, \ldots, n-1, \qquad (2)$$

lembrando que $x_i = z_i$.

VI. Resultados da Simulação em Computador

A fim de realizar a simulação em computador do esquema proposto, simulamos a BSS por meio da implementação em *software* do gerador de números pseudo-aleatórios de Park-Miller-Carta, o qual encontra-se analisado por completo em [7], [8] e [9]. A fim de testar a aleatoriedade da seqüência produzida na saída do esquema aqui proposto para realizar um codificador homofônico universal, baseado no uso de codificador diferencial e entrelaçador, empregamos a suite de testes estatísticos designados como NIST 800-22 [13], [14]. Esta suite de testes consiste de 15 testes que foram desenvolvidos para verificar o grau de aleatoriedade de uma seqüência binária longa arbitrária. Além do teste universal de Maurer, esta suite contém testes que também estão presentes em outras baterias de testes, como o *diehard*, proposto por G. Marsaglia [15]. A finalidade destes testes é identificar possíveis desvios estatísticos da aleatoriedade ideal que podem afetar uma dada seqüência binária.

Cada teste produz um valor que depende de parâmetros dos dados de entrada como, por exemplo, o comprimento de bloco da seqüência, o número de sub-blocos nos quais uma dada seqüência é dividida, etc. O valor produzido por um teste é usado para calcular um parâmetro P, que será chamado de *P-value*, o qual indica a probabilidade de que um gerador de números aleatórios perfeito teria produzido uma seqüência *menos aleatória* de que a seqüência que foi testada, dado o tipo de não-aleatoriedade investigada pelo teste.

O valor *P-value* $= 1$ indica que a seqüência, para aquele teste particular considerado, apresenta aleatoriedade perfeita. Por outro lado, o valor *P-value* $= 0$ indica que a seqüência parece ser completamente não-aleatória. Um nível de significância α, tipicamente no intervalo $[0,001 - 0,01]$, pode ser escolhido para analisar os resultados dos testes. Se o valor de *P-value* for maior que ou igual a α, considera-se que a seqüência é aparentemente aleatória. No entanto, se o valor *P-value* for menor que α, considera-se que a seqüência é aparentemente não-aleatória. Os testes estatísticos para verificar a aleatoriedade das seqüências produzidas na saída do codificador homofônico na Figura 4 utilizou o *software* disponível em [14], considerando $\alpha = 0,01$, seguindo a metodologia usada para testar os cripto-sistemas finalistas da competição para a escolha do algoritmo para representar o *Advanced Encryption Standard* [16].

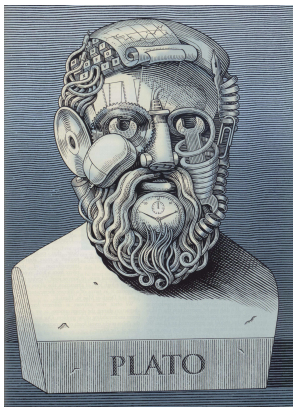

Fig. 5. Imagem de Platão (2.353×3.429 *pixels*).

TABELA I

Resultados dos testes do NIST quando a imagem de Platão é usada como fonte de informação no esquema proposto de codificação homofônica universal baseado em codificação diferencial com $n = 10$ e um entrelaçador de ordem $N = 128$.

Testes Estatísticos	(%) Aprovados
Frequency (Monobit) Test	99,47
Frequency Test within a Block	96,83
Cumulative Sums (Cusums) Test (Forward)	99,47
Cumulative Sums (Cusums) Test (Backward)	99,47
Runs Test	99,47
Longest-Run-of-Ones in a Block	98,94
Binary Matrix Rank Test	98,94
Discrete Fourier Transform (Spectral) Test	98,41
*Non-Overlapping Template Matching Test**	99
Overlapping Template Matching Test	99,47
Maurer's Universal Statistical Test	99,47
Approximate Entropy Test	100
*Random Excursions Test**	98,73
*Random Excursions Variant Test**	98,96
Serial Test (P-value 1)	100
Serial Test (P-value 2)	100
Massey's Linear Complexity Test	99,47
Porcentagem média para obter aprovação	96,83

A Tabela I apresenta resultados dos testes do NIST quando a imagem de Platão, ilustrada na Figura 5, é usada como fonte de

informação no esquema proposto de codificação homofônica universal baseado em codificação diferencial e entrelaçador de Berrou-Glavieux, com $n = 10$ e $N = 128$, em que N é a ordem do entrelaçador. Neste caso, a seqüência de *bits* produzida na saída do codificador homofônico foi aprovada em todos os testes. Lembrando que os valores indicados por um asterisco indicam a média entre todos os $P - values$ calculados pelos testes que resultam em mais de um $P-value$. A última linha da Tabela I indica a porcentagem mínima de subseqüências que precisam ser aprovadas para que a seqüência de *bits* da fonte de informação passe em cada teste.

TABELA II

RESULTADOS DOS TESTES ESTATÍSTICOS UTILIZANDO A IMAGEM DE PLATÃO COMO FONTE DE INFORMAÇÃO. OS ELEMENTOS DE CADA CÉLULA SÃO PARES (F_u, F_p), EM QUE F_u DENOTA O NÚMERO DE FALHAS NA COLUNA DE UNIFORMIDADE DOS $P - values$ E F_p DENOTA O NÚMERO DE FALHAS NA COLUNA DE PROPORÇÃO. OS PARÂMETROS n E N DENOTAM, RESPECTIVAMENTE, O COMPRIMENTO DO BLOCO CODIFICADO E A ORDEM DO ENTRELAÇADOR DE BERROU-GLAVIEUX.

n/N	8	16	32	64	128	256
2	(2,1)	(1,1)	(1,1)	(0,0)	(0,0)	(0,0)
3	(4,3)	(1,0)	(1,0)	(0,0)	(0,0)	(0,0)
4	(16,11)	(1,0)	(1,0)	(0,0)	(0,0)	(0,0)
5	(11,7)	(1,0)	(2,1)	(0,0)	(0,0)	(0,0)
6	(3,3)	(1,1)	(1,0)	(0,0)	(0,0)	(0,1)
7	(6,8)	(1,0)	(1,1)	(0,0)	(0,0)	(0,0)
8	(25,19)	(1,1)	(1,0)	(0,1)	(0,1)	(0,0)
9	(4,3)	(1,2)	(1,1)	(0,0)	(0,0)	(0,0)
10	(5,5)	(1,1)	(1,0)	(0,0)	(0,0)	(0,0)
11	(13,11)	(1,0)	(1,1)	(0,0)	(0,0)	(0,0)
12	(18,19)	(1,0)	(1,0)	(0,0)	(0,0)	(0,0)
13	(8,10)	(1,0)	(1,1)	(0,0)	(0,0)	(0,0)
14	(5,4)	(1,3)	(1,0)	(0,1)	(0,1)	(0,0)
15	(36,33)	(2,0)	(1,1)	(0,1)	(0,0)	(0,0)

Diversas simulações em computador foram realizadas considerando diferentes fontes de informação. A Tabela II apresenta alguns dos resultados obtidos, quando n se situa no intervalo $2 \leq n \leq 15$ e entrelaçadores de Berrou-Glavieux com $M = N \in \{8, 16, 32, 64, 128, 256\}$ foram usados. Utilizando um entrelaçador de ordem 64, 128 ou 256, os resultados da simulação foram muito bons para todos os valores de n considerados. Além disso, utilizando um entrelaçador de ordem 16 ou 32 este esquema ainda apresenta resultados aceitáveis em termos do número de testes nos quais foi aprovado, o que resulta em uma implementação com uma quantidade menor de memória. Em termos de desempenho o presente esquema e aquele proposto em [4] são equivalentes. Em termos de implementação o presente esquema se apresenta um pouco mais simples que aquele em [4] por não necessitar de um multiplexador, fazendo uso apenas da BSS, de um codificador diferencial e um entrelaçador, e observando que a expansão no texto-claro é bastante reduzida. Por exemplo, para o valor $n = 15$ a expansão no texto-claro fica próxima de apenas $6,7\%$, com desempenho muito bom usando $N \leq 256$.

VII. COMENTÁRIOS FINAIS

Neste artigo apresentamos um novo esquema para codificação homofônica universal, o qual combina propriedades da codificação diferencial e do entrelaçamento de símbolos. Como consequência, tanto a geração como a recuperação do texto-claro produzidos pelo esquema proposto são de fácil implementação. Testes para validação deste esquema foram realizados usando a suite de testes estatísticos NIST IR 6483 do *National Institute of Standards and Technology*. Os resultados das simulações em computador mostraram que o sistema proposto apresenta um desempenho superior aos demais com os quais foi comparado, face à suite de testes do NIST e à expansão do texto-claro, no mesmo nível daquele apresentado em [4]. Sugerimos como trabalho futuro a investigação do esquema aqui proposto quando outros tipos de entrelaçadores forem empregados.

AGRADECIMENTOS

Este trabalho recebeu apoio parcial da Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE), projeto IBPG-0222-3.04/08, e do Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq, projeto 306612/2007-0.

APÊNDICE

A. Resultados de simulações em computador do esquema proposto por Massey.

A Tabela III apresenta resultados de simulação em computador, utilizando os testes estatísticos do NIST aplicados ao esquema proposto por Massey [3], com parâmetros $m = n = 32$, usando o procedimento de codificação universal de fonte de Lynch-Davisson aplicado a blocos de comprimento 512 *bits*. Os resultados são bons apenas para o teste estatístico universal de Maurer e para o algoritmo Berlekamp-Massey, para a síntese de seqüências com registrador de deslocamento com realimentação. Nenhuma subseqüência satisfez a condição para que fosse usado o *Random Excursions Variant Test*.

B. Resultados de simulações em computador do esquema baseado na cifra de blocos descartáveis e um entrelaçador

Este apêndice apresenta resultados típicos de um teste de aleatoriedade da seqüência na saída do entrelaçador na Figura 2. A fonte de informação considerada é uma imagem de Platão no padrão RGB[1], de tamanho $26.849.646$ *bytes* (2.353×3.429 *pixels*), ilustrada na Figura 5.

A Tabela IV apresenta resultados de testes realizados usando o entrelaçador de Berrou-Glavieux com parâmetros $M = N = 64$. Os valores indicados na Tabela IV com um asterisco representam o valor médio entre todos os *P-values* computados pelo teste. A percentagem mínima exigida de blocos aprovados a fim de aprovar a seqüência de teste em cada teste, com exceção do *Random Excursions Variant Test*, é de $97,39\%$. Para o *Random Excursions Variant Test*, esta percentagem mínima exigida é de $97,04\%$. A seqüência na saída do entrelaçador na Figura 2 passou todos os testes realizados e assim, de acordo com os padrões do NIST, ela representa uma seqüência pseudo-aleatória de *bits* de alta qualidade, não apresentando desvios estatísticos.

[1]RGB é a sigla para uma das técnicas empregadas para representar uma imagem, significando *Red, Green, Blue*.

TABELA III

RESULTADOS DOS TESTES DO NIST QUANDO A IMAGEM DE PLATÃO É USADA COMO FONTE DE INFORMAÇÃO NO ESQUEMA PROPOSTO POR MASSEY, COM PARÂMETROS $n = m = 32$, USANDO O CODIFICADOR UNIVERSAL DE FONTE DE LYNCH-DAVISSON APLICADO A BLOCOS DE COMPRIMENTO $L = 512$ *bits*.

Testes Estatísticos	% Aprovados
Frequency (Monobit) Test	0
Frequency Test within a Block	0
Cumulative Sums Test (Forward)	0
Cumulative Sums Test (Backward)	0
Runs Test	0
Longest-Run-of-Ones in a Block	76,23
Binary Matrix Rank Test	0
Discrete Fourier Transform (Spectral) Test	0
*Non-Overlapping Template Matching Test**	69,07
Overlapping Template Matching Test	0
Maurer's Universal Statistical Test	92,89
Approximate Entropy Test	0
*Random Excursions Test**	0
*Random Excursions Variant Test**	0
Serial Test (P-value 1)	0
Serial Test (P-value 2)	0
Massey's Linear Complexity Test	98,53
Porcentagem média para obter aprovação	97, 52

TABELA IV

RESULTADOS DOS TESTES DO NIST QUANDO A IMAGEM DE PLATÃO É USADA COMO FONTE DE INFORMAÇÃO NO ESQUEMA DE CODIFICAÇÃO HOMOFÔNICA UNIVERSAL BASEADO EM CIFRA DE BLOCOS DESCARTÁVEIS E ENTRELAÇADOR.

Testes Estatísticos	% Aprovados
Frequency (Monobit) Test	98,25
Frequency Test within a Block	99,42
Cumulative Sums (Cusums) Test (Forward)	98,54
Cumulative Sums (Cusums) Test (Backward)	98,54
Runs Test	97,96
Longest-Run-of-Ones in a Block	98,54
Binary Matrix Rank Test	97,96
Discrete Fourier Transform (Spectral) Test	98,83
*Non-Overlapping Template Matching Test**	99,02
Overlapping Template Matching Test	98,25
Maurer's Universal Statistical Test	98,25
Approximate Entropy Test	98,54
*Random Excursions Test**	98,44
*Random Excursions Variant Test**	98,78
Serial Test (P-value 1)	98,83
Serial Test (P-value 2)	98,25
Massey's Linear Complexity Test	98,83
Porcentagem média para obter aprovação	97,39

C. O entrelaçador de Berrou-Glavieux

O entrelaçador proposto por Berrou e Glavieux [11], [12] é um entrelaçador de bloco e consiste de uma matriz $N \times M$, na qual $N = 2^l$ e $M = 2^m$, com $l \geq 3$ e $m \geq 3$, tal que sua saída no instante de tempo i, denotada por y_i, é idêntica à sua entrada no instante de tempo $\pi(i)$, denotada como $x_{\pi(i)}$, para $0 \leq i < N \cdot M = T$. Berrou e Glavieux propuseram para uso neste entrelaçador os números primos $p(j)$, $0 \leq j \leq 7$, como mostrados na Tabela V.

A função de permutação $\pi(i)$ foi definida do seguinte modo.

$$\pi(i) = c(i) + M \cdot r(i), \tag{3}$$

TABELA V

NÚMEROS PRIMOS USADOS NO ENTRELAÇADOR DE BERROU-GLAVIEUX.

j	0	1	2	3	4	5	6	7
p(j)	17	37	19	29	41	23	13	7

em que

$$r(i) = \left(\frac{M}{2} + 1\right)(r_0 + c_0) \pmod{M}$$

$$c(i) = p(s)(c_0 + 1) - 1 \pmod{N}$$

$$s = r_0 + c_0 \pmod{8}$$

$$r_0 = i \pmod{M}$$

$$c_0 = \frac{i - r_0}{M}$$

REFERÊNCIAS

[1] D. R. Simões e V. C. da Rocha Jr., "Um esquema de codificação homofônica universal utilizando o algoritmo LZW", XXVI Simpósio Brasileiro de Telecomunicações, Rio de Janeiro, pags. 1-6, 2 a 5 de setembro de 2008.

[2] C. Günther, "A Universal Algorithm for Homophonic Coding", *Advances in Cryptology - Eurocypt '88*, (Ed. C. G. Günther) LNCS no 330, Springer-Verlag, pp. 405-41, 1988.

[3] J. L. Massey, "Some Applications of Source Coding in Criptography", *European Transactions on Telecommunications*, vol.5, no. 4, pp. 7/421-15/429, 1994.

[4] D. R. Simões and V. C. da Rocha Jr., "A versatile universal homophonic coding scheme", Tenth International Symposium on Communication Theory and Applications (ISCTA '09), Ambleside, Lake District, UK, pp 1-4, 13th - 17th July, 2009.

[5] H. K. Jendal, Y. J. B. Kuhn and J. L. Massey, "An Information-Theoretic Approach to Homophonic Substitution", *Advances in Cryptology - Eurocypt '89*, (Eds. J.-J. Quisquater e J.Vanderwalle), Lect. Notes in Comp. Sci., No. 434, Springer, pp 382-394, 1990.

[6] C. E. Shannon, "Communication Theory of Secrecy Systems", *Bell Sys. Tech. J.*, Vol. 28, pp. 656-715, Outubro 1949.

[7] Stephen K. Park and Keith W. Miller, "Random Number Generators: Good ones are Hard to Find", *Communications of ACM*, Vol. 31, No. 10, Outubro 1988.

[8] David G. Carta, "Two Fast Implementations of the "Minimal Standard"Random Number Generator", *Communications of ACM*, Vol. 33, No. 1, Janeiro 1990.

[9] G. S. Fishman and L. R. Moore, "An Exhaustive Analysis of Multiplicative Congruential Random Number Generators With Modulus $2^{31} - 1$", *SIAM J. Sci. Stat. Comput.*, vol. 7, Janeiro de 1986, pp. 24-45.

[10] C. Heegard and S. B. Wicker, *Turbo Codes*, Kluwer Academic Publishers, Boston, 1999.

[11] C. Berrou and A. Glavieux, "Near Optimum Error Correcting Coding and Decoding: Turbo-Codes", *IEEE Transactions on Communications*, COM-44(10), pp. 1261-1271, 1996.

[12] C. Berrou, A. Glavieux and P. Thitimajshima, "Near Shannon Limit Error-Correcting Coding and Decoding: Turbo Codes", *International Conference on Communications Conference*, pp. 1064-1070, 1993.

[13] "A statistical test suite for random and pseudorandom number generators for cryptographic applications", National Institute of Standards and Technology, Special Publication 800-22, 2001, available on-line: *http://csrc.nist.gov/rng/SP800-22b.pdf*, *http://csrc.nist.gov/publications/nistpubs/800-22-rev1/SP800-22rev1.pdf*.

[14] "NIST statistical test suite (version 1.8)", March 2005, available online: *http://csrc.nist.gov/rng/*.

[15] "The Marsaglia Random Number CDROM including the Diehard Battery of Tests of Randomness", available on-line: *http://www.stat.fsu.edu/pub/diehard/*.

[16] J. Soto and L. Bassham, "Randomness Testing of the Advanced Encyption Standard Finalist Candidates", NIST IR 6483, National Institute of Standards and Technology, Gaithersburg, 2000.