Mecanismo identificador para n-cifras de bloco

William A. R. de Souza[*], Luís Alfredo Vidal de Carvalho[*], José A. M. Xexéo[†]
[*]COPPE/UFRJ – Universidade Federal do Rio de Janeiro
Caixa Postal 68511 – CEP 21945-970 - Rio de Janeiro – RJ – Brasil
[†]Seção de Engenharia de Sistemas – Instituto Militar de Engenharia
Pça. General Tibúrcio 80, Praia Vermelha, CEP 22290-270 – Rio de Janeiro – RJ
Email: {william, alfredo}@cos.ufrj.br, xexeo@ime.eb.br

***Abstract*—This paper proposes a mechanism to identify n-block ciphers in ECB and CBC mode of operation, from patterns found on a set of cryptogams. This result demonstrates the existence of intrinsic properties in mathematical models of the ciphers, which create a signature in cryptograms. Based on unsupervised categorization technique, experiments are performed on a set of cryptogams which were enciphered by the finalist algorithms of AES contest: MARS, RC6, Rijndael, Serpent and Twofish (particular case of n = 5); with 128-bit key. The processes of clustering and categorization used were successful, allowing 100% of precision in ECB mode. The CBC mode reached 100% of precision in clustering, although it failed in the categorization process.**

I. INTRODUÇÃO

A verificação da robustez de algoritmos criptográficos[1] exige o teste de requisitos de segurança. Estes requisitos se baseiam na resistência contra diferentes tipos de ataques, alguns disponíveis na literatura científica. Alguns desses ataques podem ser vistos em [1]. Dentre eles, estão os ataques de Criptoanálise Linear [2] e Diferencial [3]. Apesar da maioria dos ataques não serem práticos em função dos atuais recursos computacionais, eles têm sido úteis para definir a robustez criptográfica de criptossistemas simétricos, como, por exemplo, as cifras de blocos contemporâneas.

Na essência, todos esses ataques, mesmo não sendo práticos, visam obter propriedades estatísticas, usando pares de texto claro e texto cifrado (criptograma), independente da chave criptográfica, que apóiem a verificação da robustez criptográfica das cifras de blocos. O NIST[2] propõe uma bateria de testes estatísticos [4] no intuito de verificar se um determinado algoritmo criptográfico pode ser utilizado como gerador de números pseudo-aleatórios. A aprovação nestes testes estatísticos é uma condição necessária, mas não suficiente, para se caracterizar a robustez criptográfica desses algoritmos. Murphy [5], por exemplo, contestou sobre a adequabilidade destes testes, indicando que os mesmos precisariam de testes complementares.

Visando, ainda, a busca de fraquezas em um algoritmo criptográfico pelo emprego de técnicas estatísticas, pode ser citado o "Ataque de Distinção" [6], que aplica o teste χ^2 em um conjunto de criptogramas gerados pelo RC6. A conclusão desse trabalho indica que o RC6 é vulnerável a tal ataque com 15 iterações e complexidade da ordem de 2^{125}. Os autores consideram em aberto esse tipo de ataque ao RC6 com 16 iterações. Com relação ao RC6, uma contramedida a esse tipo de ataque foi proposta por Ueda e Terada [7] com intuito de fortalecer este algoritmo.

Algumas dessas técnicas foram usadas com sucesso para agrupar criptogramas, no modo ECB, em função da chave, com técnicas de recuperação de informações aplicadas aos algoritmos DES, AES e RSA [8], [9] e [10]. Outras técnicas foram usadas para identificar cifras, no modo ECB, com máquinas de vetor de suporte aplicadas aos algoritmos DES, Triple DES, Blowfish, AES e RC5 [11] e com métodos de histograma e de predição de bloco aplicados aos algoritmos DES, Triple DES, Blowfish, RC5 e AES [12].

Esses trabalhos citados, entretanto, estão limitados a identificar e distinguir cifras com mecanismos especializados em um único algoritmo, o qual se pretende identificar. Assim, nesses trabalhos, dado um conjunto de criptogramas, o mecanismo identificador pode separar do conjunto de criptogramas apenas os criptogramas gerados pelo algoritmo no qual é especializado. Além disso, não há relatos de sucesso utilizando o modo CBC.

Este trabalho contorna as limitações acima, ao propor um mecanismo identificador para n-cifras de blocos. O mecanismo será aplicado a um caso particular de n = 5 onde são utilizadas as cifras de blocos: MARS, RC6, Rinjdael, Serpent e Twofish; nos modos de operação ECB e CBC. O método é baseado em técnicas de classificação de textos e demonstra que o sucesso se deve à correta separação dos criptogramas em grupos e também a existência de propriedades intrínsecas nos modelos matemáticos dessas cifras, as quais criam uma assinatura nos criptogramas.

No modo ECB, o agrupamento[3] foi bem sucedido, ocorrendo sempre a formação de cinco grupos, um para cada cifra, e cada grupo contendo apenas criptogramas gerados com a mesma cifra, o que permitiu a identificação correta na fase de classificação. No modo CBC o agrupamento foi realizado

[1] Neste texto, os termos algoritmo, algoritmo criptográfico e cifra são usados indistintamente.

[2] National Institute of Standard and Tecnology.

[3] Neste texto, os termos agrupamento e separação em grupos são usados indistintamente.

com 100% de precisão, mas falhou no processo de classificação.

O trabalho está organizado como segue. Na seção 2 são apresentadas a metodologia e definições deste trabalho. O procedimento utilizado é descrito na seção 3. Os experimentos, os resultados obtidos e as avaliações compõem a seção 4. Na seção 5 e realizada uma discussão sobre os resultados obtidos, considerando os experimentos e as avaliações. A conclusão é apresentada na seção 6

II. METODOLOGIA E DEFINIÇÕES UTILIZADAS

A. *Criptografia*

Um sistema criptográfico pode ser definido como uma quíntupla (T,C,K,E,D), onde as seguintes condições são satisfeitas [13]:

1. T é um conjunto finito de possíveis textos claros;
2. C é um conjunto finito de possíveis criptogramas;
3. K, o espaço de chaves, é um conjunto finito de possíveis chaves;
4. Para cada $k \in K$, existe uma regra de cifração $e_k \in E$ e uma correspondente regra de decifração $d_k \in D$. $e_k : T \to C$ e $d_k : C \to T$ são funções, tais que $d_k(e_k(t)) = t$ para todo texto claro $t \in T$.

B. *Cifras de Bloco*

Sejam c um criptograma, t um texto claro e e_k uma função de cifração, modificada por uma chave k. $e_k : t \to c$, dividindo t em $t_1,t_2,...,t_n$, de tal forma que $|t_1| = |t_2| = ... = |t_n|$ e $|t_1| + |t_2| + ... + |t_n| = |t|$, e transformando $t_1,t_2,...,t_n$, um a um, em $c_1,c_2,...,c_n$, de tal forma que $|c_1| = |c_2| = ... = |c_n|$, $|c_1| + |c_2| + ... + |c_n| = |c|$ e $|c| = |t|$. Os tamanhos de $t_1,t_2,...,t_n$ dependem do algoritmo e, muitas vezes, do tamanho da chave utilizada.

C. *Composição de Cifras de Bloco*

Sejam c um criptograma, t um texto claro, T um conjunto finito de possíveis textos claros, C_1 e C_2 dois conjuntos finitos de possíveis criptogramas e $e_k^1 : T \to C_1$ e $e_k^2 : C_1 \to C_2$ duas funções. A composição de e_k^2 com e_k^1, denotada por $e_k^2 \circ e_k^1$ é uma função de T para C_2, tal que $e_k^2 \circ e_k^1 : T \to C_2$, e definida por $(e_k^2 \circ e_k^1)(t) = e_k^2(e_k^1(t))$, $\forall t \in T$.

A definição acima pode ser estendida para três ou mais funções. Assim, para as funções e_k^1, e_k^2,..., e_k^m, pode-se definir $e_k^m \circ ... \circ e_k^2 \circ e_k^1$, contanto que o domínio de e_k^m seja igual ao contradomínio de e_k^{m-1} e assim por diante [1].

D. *Modos de Operação*

Seja c_i um bloco de um criptograma. Definimos como ECB, o modo de operação onde $\forall i = 1,...,n$, $c_i = e_k(t_i)$, onde t_i é um bloco de texto claro. Alguns autores não recomendam o uso deste modo de operação, como [14]. Entretanto, este é o modo utilizado nos testes de aleatoriedade do NIST [15], por ser o que permite que a cifração seja a mais semelhante com a transformação original proposta pela cifra.

Seja c_i um bloco de um criptograma. Definimos como CBC, o modo de operação onde $\forall i = 1,...,k$, $c_i = e_k(t_i \oplus c_{i-1})$, onde t_i é um bloco de texto claro. Nota-se que tal modo faz um encadeamento entre os blocos, o que confere maior confusão ao criptograma gerado. O bloco c_0 é definido por um Vetor de Inicialização (IV). Este modo é amplamente utilizado em criptossistemas [14].

E. *Medida de Similaridade*

A coleção de criptogramas é modelada em um espaço de vetores. Desta forma, para representar os criptogramas são utilizados vetores de dimensão-*n*, onde *n* é o número de blocos distintos em todo o conjunto de criptogramas. Assim, sejam os vetores $c_i = (c_{1,i},c_{2,i},...,c_{n,i})$ e $c_j = (c_{1,j},c_{2,j},...,c_{n,j})$, dois criptogramas para os quais se deseja obter a similaridade. O valor relacionado à $c_{k,i}$, onde *k* representa o *k*-ésimo bloco de c_i, é a freqüência do bloco *k* em c_i. O valor relacionado à $c_{k,j}$, onde *k* representa o *k*-ésimo bloco de c_j, é a freqüência do bloco k em c_j.

A similaridade entre dois criptogramas c_i e c_j foi associada ao co-seno do ângulo e calculada pela fórmula (1), tal que $0 \le S \le 1$. Quanto maior o valor de *S*, maior a similaridade entre os criptogramas. O ângulo do co-seno é amplamente utilizado em tarefas de Recuperação de Informações [16]. Então, cria-se uma matriz de similaridades, armazenando, em suas células, os valores de similaridade dos pares de criptogramas da coleção.

$$S_{Co-seno}(c_i,c_j) = \frac{\sum_{k=1}^{n}(c_{i,k} \times c_{j,k})}{\sqrt{\sum_{k=1}^{n}(c_{i,k})^2 \times \sum_{k=1}^{n}(c_{j,k})^2}} \quad (1)$$

F. *Agrupamento*

O método de agrupamento deste trabalho utiliza a técnica hierárquica aglomerativa da ligação simples [17] formando grupos dispersos [18], o que permite que dois criptogramas quaisquer que estejam em um grupo, possuam valor de similaridade mais baixo que a similaridade do próprio grupo. Assim, com a conclusão do procedimento, *n* criptogramas são categorizados em *m* grupos. O valor de *m* é desconhecido no início do procedimento.

Inicialmente, cada criptograma pertencerá a um único grupo. A seguir, identifica-se, na matriz de similaridades, o par de criptogramas com o maior valor de similaridade para formar o primeiro grupo. Atribui-se ao grupo um valor de "similaridade de grupo" igual ao maior valor de similaridade existente entre os pares de criptogramas pertencentes ao grupo.

A matriz de similaridade é atualizada pela substituição da similaridade do par pela similaridade do grupo. Esse procedimento é repetido até que um critério de parada seja alcançado.

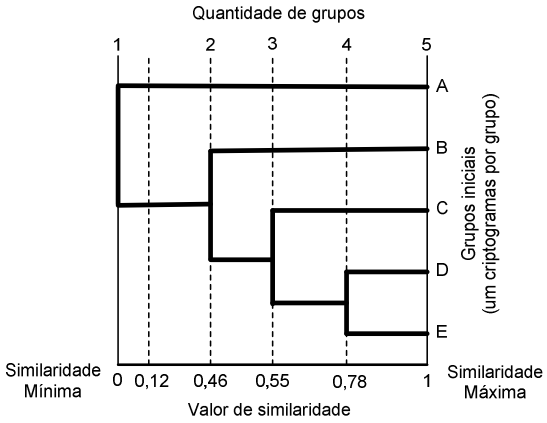

Fig. 1. Dendrograma.

A estrutura final do agrupamento, representando a ordem em que as inclusões e junções dos grupos ocorrem é representada por um dendrograma [17] (figura 1). Observando essa figura, pode-se notar que os grupos foram formados a partir de um determinado valor de similaridade, o qual pode ser utilizado como o critério de parada citado anteriormente.

Neste trabalho, utiliza-se um valor de similaridade próximo de zero como critério de parada, pois é pouco provável a repetição de blocos ao longo dos criptogramas.

G. Avaliação dos Grupos

Na avaliação da qualidade do agrupamento foram utilizadas as medidas revocação e precisão [19] e [20] (Figura 2).

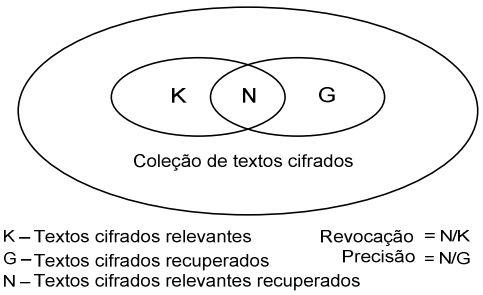

Fig. 2. Revocação e Precisão.

Os valores de revocação e precisão são definidos como segue. Suponha que K seja o conjunto formado pelos criptogramas cifrados com um determinado algoritmo Δ. Seja G o agrupamento construído pelo método proposto e que supostamente contém os criptogramas gerados pelo algoritmo Δ. Seja $|k|$ o número de elementos no conjunto K, $|g|$ o número de elementos de G e n o número de elementos do conjunto K presentes no grupo G. Então, os valores de Revocação e Precisão são obtidos pelas fórmulas 2 e 3, respectivamente, tal que $0 \le R \le 1$ e $0 \le P \le 1$. Revocação, portanto, indica a capacidade do método de recuperar todos os criptogramas relevantes[4]. Precisão, por sua vez, indica a capacidade do método de recuperar apenas criptogramas relevantes.

$$R = \frac{n}{|k|} \tag{2}$$

$$P = \frac{n}{|g|} \tag{3}$$

III. Descrição do Mecanismo Proposto

O mecanismo proposto considera os blocos dos criptogramas como palavras de um texto de idioma desconhecido [21] e agrupa os criptogramas com base na similaridade existente entre eles, similaridade essa medida com base na freqüência com que palavras (blocos) ocorrem em cada um dos textos (criptogramas). O procedimento é não supervisionado, ou seja, os processos são executados sem o conhecimento dos textos claros, dos algoritmos criptográficos, da chave utilizada no processo de cifração ou da quantidade de grupos que serão formados.

Desta forma, os criptogramas são representados de maneira que seja possível determinar similaridades entre eles, dois a dois [17], e construir uma matriz de similaridades. A partir dessa matriz é realizado o agrupamento pelo método da ligação simples, descrito na seção anterior. Então, define-se um valor de similaridade para análise de grupos, que determina a quantidade de grupos a ser formada. Por fim, a qualidade do agrupamento é avaliada pelas medidas revocação e precisão (Figura 3).

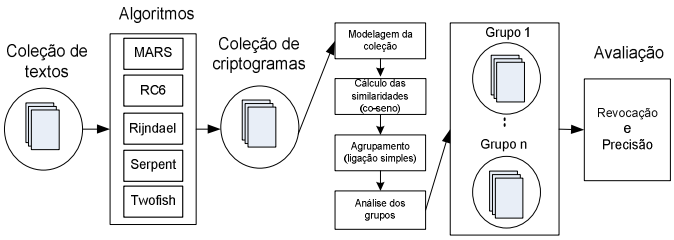

Fig. 3. Esquema do processo de agrupamento.

A seguir, a pertinência de cada novo criptograma a um dos grupos anteriormente formados (classificação) indica o algoritmo que cifrou o mesmo ou indica que uma nova cifra, diferente das anteriormente classificadas, foi detectada (figura 4). Tal procedimento pode ser considerado, por analogia, como um ataque de distinção.

Na figura 4, nota-se que o canal de informação é vulnerável a ataques passivos, como análise de tráfego, mesmo quando protegido por criptografia. Realizando a análise de tráfego, um atacante poderia obter informações sobre alguns padrões de tráfego relativos às mensagens, como freqüência e tamanho da

[4] Criptogramas relevantes são aqueles que pertencem naturalmente a um determinado grupo. Por exemplo, em um grupo formado por criptogramas cifrados pelo MARS, os criptogramas relevantes são os cifrados pelo MARS. Os demais criptogramas, cifrados por outros algoritmos, não são relevantes.

mensagem. Com o mecanismo proposto, pode-se obter informação mais útil como a identificação do algoritmo criptográfico ou a detecção de mudança do algoritmo utilizado na cifração.

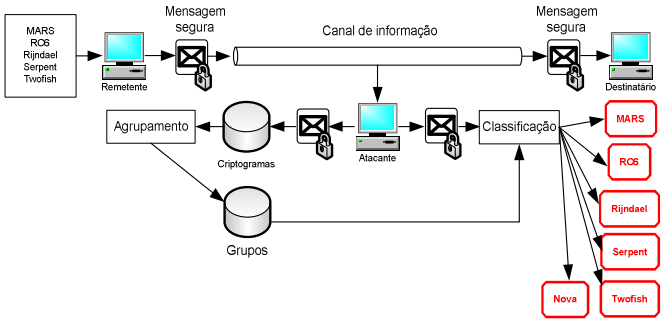

Fig. 4. Modelo para a identificação de cifra.

O mecanismo pode ser utilizado também para detectar a mudança na chave criptográfica utilizada para cifrar a mensagem que trafega na rede.

IV. Experimentos, Resultados e Avaliações

Os textos claros foram obtidos em [22] e os resultados gerados pela ferramenta WARS Text[5], desenvolvida em [9][10] e [21]. Todos os experimentos utilizaram os mesmos 30 textos claros para cada uma de nove coleções, com tamanhos de: 1024, 1536, 2048, 2560, 3072, 4096, 6144, 8192 e 10240 bytes. Todos os experimentos foram realizados com blocos de 128 bits.

A. *Experimento 1: Identificação de Cifra*

O objetivo do experimento é: a) Fase 1: separar os criptogramas em cinco grupos de maneira que cada grupo contenha criptogramas gerados pela mesma cifra e somente por ela; b) Fase 2: Classificar cada criptograma por cifra de bloco, de acordo com os grupos criados na fase 1. O resultado das duas fases identifica a cifra de bloco.

Para cada modo de operação (ECB e CBC) são definidas nove coleções, cada uma com determinado tamanho. Cada coleção é composta por 150 criptogramas, gerados por cinco algoritmos: MARS, RC6, Rinjdael, Serpent e Twofish, a partir dos mesmos 30 textos claro e uma única chave de 128 bits para cada modo, portanto, o experimento utiliza 1350 criptogramas (150 criptogramas x 9 coleções = 1350). Os primeiros 70% dos criptogramas de cada coleção foram usados na fase 1 e os 30% restantes foram usados na fase 2.

Resultado da Fase 1: Separação dos Grupos

Na tabela 1, observa-se que o procedimento separou corretamente os criptogramas com tamanho a partir de 1024 bytes, de maneira que cada grupo continha criptogramas gerados pelo mesmo algoritmo e somente por ele, obtendo valor máximo de precisão nos dois modos de operação. A discriminação é perfeita e, mesmo no modo CBC, o procedimento não mistura no mesmo grupo criptogramas oriundos de algoritmos diferentes.

TABELA 1
RESULTADO DA SEPARAÇÃO EM GRUPOS

Tamanho dos criptogramas (em bytes)	ECB		CBC	
	Precisão	Revocação	Precisão	Revocação
1024	1	0,17	1	0,03
1536	1	0,20	1	0,03
2048	1	0,33	1	0,03
2560	1	0,50	1	0,03
3072	1	0,87	1	0,03
4096	1	0,97	1	0,03
6144	1	1	1	0,03
8192	1	1	1	0,03
10240	1	1	1	0,03

Os resultados com a métrica revocação não são tão bons. O modo ECB, para alguns tamanhos de criptogramas, não alcançou o valor máximo, o que é de se esperar visto que textos com menos de 4096 bytes ainda geram dicionário muito pequeno. No modo CBC, os valores de revocação mostram que o procedimento terminou com a quantidade grupos inicial, onde cada grupo continha apenas um criptograma, indicando que não foram achadas similaridades entre os criptogramas, o que confirma o melhor nível de aleatoriedade dado pelo modo CBC.

Resultado da Fase 2: Classificação das Cifras

Nesta fase o experimento utiliza, para cada modo, 405 criptogramas (30% da coleção), submetidos um a um ao processo de classificação, com a finalidade de alocar cada criptograma a um dos grupos anteriormente formados e, assim, identificar a cifra.

Observa-se na tabela 2 que no modo ECB o experimento obteve sucesso identificando corretamente os algoritmos geradores de 303 (75%) criptogramas submetidos ao processo. Nota-se que para os experimentos cujo total de acertos foi menor do que 100%, um novo grupo foi criado, o que garante a precisão máxima.

TABELA 2
RESULTADO DA IDENTIFICAÇÃO DAS CIFRAS DE BLOCO

Tamanho dos criptogramas (em bytes)	ECB		CBC	
	Acertos	Criação de novo grupo	Acertos	Criação de novo grupo
1024	22 %	Sim	0 %	Sim
1536	24 %	Sim	0 %	Sim
2048	69 %	Sim	0 %	Sim
2560	91 %	Sim	0 %	Sim
3072	89 %	Sim	0 %	Sim
4096	78 %	Sim	0 %	Sim
6144	100 %	Não	0 %	Sim
8192	100 %	Não	0 %	Sim
10240	100 %	Não	0 %	Sim

[5] A ferramenta, a coleção de criptogramas e os textos claros utilizados estão disponíveis por meio de contato com os autores.

Já no modo CBC, o experimento não foi capaz de classificar os criptogramas, dado que não se pôde obter similaridade entre os mesmos. Contudo, foram criados novos grupos, o que indica que criptogramas cifrados por algoritmos diferentes não se misturam em um mesmo grupo, garantindo a máxima precisão.

B. Experimento 2: Separação em Grupos Usando Composição de 5 Cifras no Modo ECB

Dada a evidência de que as propriedades intrínsecas dos modelos matemáticos de cada cifra geram assinaturas nos criptogramas, o que foi parcialmente demonstrado pelo experimento 1, o objetivo deste experimento é verificar se a composição de cifras diferentes permitem gerar assinaturas nos criptogramas, de tal forma que a última cifra utilizada na composição seja corretamente agrupada.

Cada coleção é composta por 150 criptogramas cifrados com uma única chave aleatória k de 128 bits, por composições com cinco cifras de blocos, conforme abaixo:

$MARS_k^{ECB}$ ($RC6_k^{ECB}$ ($Rijndael_k^{ECB}$ ($Serpent_k^{ECB}$ ($Twofish_k^{ECB}$ ($texto_1$,...,$texto_{30}$)))));

$RC6_k^{ECB}$ ($Rijndael_k^{ECB}$ ($Serpent_k^{ECB}$ ($Twofish_k^{ECB}$ ($MARS_k^{ECB}$ ($texto_1$,...,$texto_{30}$)))));

$Rijndael_k^{ECB}$ ($Serpent_k^{ECB}$ ($Twofish_k^{ECB}$ ($MARS_k^{ECB}$ ($RC6_k^{ECB}$ ($texto_1$,...,$texto_{30}$)))));

$Serpent_k^{ECB}$ ($Twofish_k^{ECB}$ ($MARS_k^{ECB}$ ($RC6_k^{ECB}$ ($Rijndael_k^{ECB}$ ($texto_1$,...,$texto_{30}$)))));

$Twofish_k^{ECB}$ ($MARS_k^{ECB}$ ($RC6_k^{ECB}$ ($Rijndael_k^{ECB}$ ($Serpent_k^{ECB}$ ($texto_1$,...,$texto_{30}$))))).

Os resultados na tabela 3 mostram que o mecanismo separou corretamente os criptogramas considerando a última cifra utilizada na composição, obtendo precisão máxima. Isto indica que não ocorreu o caso de criptogramas gerados com cifras diferentes se juntarem no mesmo grupo. Assim, percebe-se que as assinaturas geradas pelas cifrações anteriores na composição são eliminadas e apenas a assinatura da última cifração permanece, permitindo a separação correta. Os valores de revocação, para alguns tamanhos de criptogramas, não alcançaram o valor máximo.

Visto que os resultados da separação em grupos neste experimento foram idênticos ao experimento 1 (tabela 1 – ECB), decorre que a classificação também será idêntica.

TABELA 3
SEPARAÇÃO EM GRUPOS USANDO COMPOSIÇÃO DE 5 CIFRAS DE BLOCO

Tamanho dos criptogramas (em bytes)	Precisão	Revocação
1024	1	0,17
1536	1	0,20
2048	1	0,33
2560	1	0,50
3072	1	0,87
4096	1	0,97
6144	1	1
8192	1	1
10240	1	1

C. Experimento 3: Separação em Grupos Usando Composição de 5 Cifras nos Modos ECB e CBC

Observa-se no experimento 1 que embora a precisão tenha sido máxima, tanto a separação quanto a classificação não tiveram sucesso no modo CBC. Considerando os resultados dos experimentos 1 e 2, no experimento 3, objetiva-se testar: a) dada uma composição com cinco cifras de blocos, se a mistura gerada pelas quatro primeiras cifrações usando o modo CBC pode ser comprometida por uma última cifração no modo ECB; e b) dada uma composição com cinco cifras de blocos, se a repetição de blocos gerada por quatro cifrações no modo ECB pode ser desfeita por uma última cifração no modo CBC.

Para cada um dos objetivos "a" e "b" acima, são definidas nove coleções, cada uma com determinado tamanho. Cada coleção é composta por 150 criptogramas, cifrados com uma única chave aleatória k de 128 bits, por composições com cinco cifras de blocos, semelhante ao experimento 2, onde para o objetivo "a" as quatro primeiras cifras usam o modo CBC e a última usa o modo ECB e para o objetivo "b" as quatro primeiras cifras usam o modo ECB e a última usa o modo CBC.

TABELA 4
QUATRO COMPOSIÇÕES NO MODO CBC E A ÚLTIMA NO MODO ECB

Tamanho dos criptogramas (em bytes)	Quatro cifrações no modo CBC		Última cifração no modo ECB	
	Precisão	Revocação	Precisão	Revocação
1024	1	0,03	1	0,03
1536	1	0,03	1	0,03
2048	1	0,03	1	0,03
2560	1	0,03	1	0,03
3072	1	0,03	1	0,03
4096	1	0,03	1	0,03
6144	1	0,03	1	0,03
8192	1	0,03	1	0,03
10240	1	0,03	1	0,03

Na tabela 4, pode-se ver que, como esperado, o resultado até a quarta cifração não gerou repetição de nenhum padrão e a quinta cifração no modo ECB não comprometeu o efeito das cifrações anteriores.

TABELA 5
QUATRO COMPOSIÇÕES NO MODO ECB E A ÚLTIMA NO MODO CBC

Tamanho dos criptogramas (em bytes)	Quatro cifrações no modo CBC		Última cifração no modo ECB	
	Precisão	Revocação	Precisão	Revocação
1024	1	0,17	1	0,03
1536	1	0,20	1	0,03
2048	1	0,33	1	0,03
2560	1	0,50	1	0,03
3072	1	0,87	1	0,03
4096	1	0,97	1	0,03
6144	1	1	1	0,03
8192	1	1	1	0,03
10240	1	1	1	0,03

Os resultados na tabela 5 mostram que uma última cifração no modo CBC destrói todos os padrões repetidos e que se propagaram ao longo das quatro cifrações anteriores no modo ECB. Os valores de precisão e revocação se mantiveram coerentes com os experimentos anteriores.

V. DISCUSSÃO

Analisando os resultados dos experimentos 1, 2 e 3, no modo ECB, nota-se que os valores de precisão e revocação se mantiveram constantes. Isso se deve ao padrão de repetição dos blocos ao longo da coleção de criptogramas.

Dos experimentos, percebe-se que a identificação das cifras é possível por meio de um processo de classificação, o qual tem seu percentual de acerto relacionado ao processo de separação em grupos: quanto melhor a separação, maior a taxa de acerto na identificação. As melhores separações ocorrem para os maiores tamanhos de criptogramas, uma vez que os blocos se repetem mais nos tamanhos maiores.

Demonstra-se com o experimento 2 que a identificação correta das cifras é conseqüência também das propriedades intrínsecas dos modelos matemáticos das cifras, uma vez que cada uma delas transforma as suas entradas em criptogramas de uma maneira particular.

O experimento 3, indica que em uma composição de cifras de blocos: a) quatro cifrações no modo CBC não deixam assinaturas que possam ser usadas por uma última cifração no modo ECB; e b) quatro cifrações no modo ECB deixam assinaturas, mas estas são destruídas por uma última cifração no modo CBC.

Sugerem-se como trabalhos futuros:

1) Formalizar o mecanismo proposto para n-cifras de blocos;
2) Explorar característica do modo de operação OFB para o caso particular discutido neste trabalho, dado que o estudo deste modo sugere um potencial de classificação melhor do que o modo CBC; e
3) Realizar experimentos semelhantes aos deste trabalho, para cifras de bloco que apresentam estruturas algébricas já identificadas, como: RSA (anel), Curvas Elípticas (grupo aditivo) e ElGamal (grupo multiplicativo), visando validar o mecanismo proposto para cifras que apresentam modelos matemáticos isomorfos. Como exemplo, um dos objetivos nesse caso seria identificar algoritmos a partir de criptogramas gerados por duas curvas elípticas isomórficas.

VI. CONCLUSÃO

O trabalho propõe um mecanismo de identificação para n-cifras de blocos e comprova o seu sucesso com um caso particular de n = 5, identificando corretamente as cifras de blocos MARS, RC6, Rijndael, Serpent e Twofish a partir de padrões detectados sobre um conjunto de criptogramas cifrados por elas.

Os experimentos demonstram que a correta identificação dos algoritmos depende da correta separação dos criptogramas em grupos de algoritmos, onde o sucesso da separação é dado pelas medidas precisão e revocação. Demonstram, também, que a identificação correta das cifras é influenciada pelas propriedades intrínsecas dos modelos matemáticos das cifras, dado que cada uma realiza as suas transformações de uma maneira particular.

Concluí-se que uma coleção de criptogramas no modo ECB, pode ser corretamente identificada com 100% de acerto, desde que tais criptogramas tenham pelo menos 6144 bytes.

As maiores contribuições deste trabalho são a proposta de um mecanismo generalizado para a identificação de cifras e a demonstração da influência de propriedades intrínsecas dos modelos matemáticos destas cifras na identificação correta das mesmas.

REFERÊNCIAS

[1] Menezes, A. J, Oorschot, P. C. Van e Vanstone, S. A, (1996), Handbook of applied cryptography, CRC Press.

[2] Matsui, M. (1993), Linear cryptanalysis method for DES cipher. In: Eurocrypt 1993, volume 765, LNCS, Journal of Cryptology, 386 – 397. Springer-Verlag.

[3] Biham, E. e Shamir A. (1991) Differential cryptanalysis of DES-like cryptsystems. Journal of Cryptology, 4(1): 3 – 72.

[4] NIST (2008). A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications. NIST Special Publication 800-22. Revision 1. Washington D.C.

[5] Murphy, Sean (2000). "The Power of NIST's Statistical Testing of AES Candidates. Information Security Group, Royal Holloway, University of London.

[6] Knudsen, L.R. and Meier, W. (2000). Correlations in RC6 with a Reduced Number of Rounds. Proceedings of the 7th International Workshop on Fast Software Encryption.

[7] Ueda, T. K. e Terada, R. (2007). Uma Versão Mais Forte do Algoritmo RC6 contra a criptoanálise χ^2. VII Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais.

[8] Carvalho, C. A. B. de (2006). O uso de técnicas de recuperação de informações em criptoanálise. . Dissertação de Mestrado – Instituto Militar de Engenharia. Disponível em: http://www.ime.eb.br

[9] Souza, W. A. R. de (2007). Identificação de padrões em criptogramas usando técnicas de classificação de textos. Dissertação de Mestrado – Instituto Militar de Engenharia. Disponível em: http://www.cos.ufrj.br/~william/teseIME.pdf

[10] Souza, W. A. R; Xexéo, J. A. M; Oliveira, C. M. G. M. (2008). Método de Agrupamento de Criptogramas em Função das Chaves de Cifrar. IV Workshop em Algoritmos e Aplicações de Mineração de Dados (SBBD/SBES).

[11] Dileep, A. D. e Sekhar, C. C. (2006), "Identification of block ciphers using support vector machines," in International Joint Conference on Neural Networks (IJCNN 2006), Vancouver, BC, Canada, July 2006, pp. 2696-2701

[12] Nagireddy, S. (2008). A Pattern Recognition Approach to Block Cipher Identification. Master of Science Dissertation – Indian Institute of Technology Madras. Disponível em: http://lantana.tenet.res.in/website_files/thesis/MS/sreenivasuluNR_thesis.pdf

[13] Stinson, D. R. Cryptography: theory and practice. 3th ed. Boca Raton: CRC, 2006.

[14] Ferguson, N., Schneier, B. e Kohno, Tadayoshi (2010), Cryptography Engineering: design principles and practical applications, Wiley.

[15] Soto, J. (1999). Randomness Testing of the Advanced Encryption Standard Candidates Algorithms. NIST Internal Report. NIST IR 6390. Disponível em: http://csrc.nist.gov/publications/nistir/ir6390.pdf [capturado 10 ago. 2008].

[16] Harman, D. (1992), "Ranking algorithms". In Information retrieval: data structures and algorithms, Edited by William Frakes and Ricardo Yates, Prentice Hall, p. 363–392.

[17] Rasmussen, E. (1992), “Clustering algorithms”. In Information retrieval: data structures and algorithms, Edited by W. Frakes and R. Yates, Prentice Hall, p. 419-442.

[18] Jain, A. K, Murty, M. N e Flynn, P. J (1999). Data Clustering: A Review. ACM Computing Surveys, Vol. 31, No 3, Setember 1999. p. 264-323.

[19] Yates, R.B. e Neto, B. R. (1999), Modern information retrieval. Addison Wesley.

[20] Fung, B. C. M., Wang, K. e Ester M. (2003), Hierarchical document clustering using frequent itemsets. Proceedings of the SIAM International Conference on Data Mining, San Francisco.

[21] Souza, W. A. R; Carvalho, L. A. V; Xexéo, J. A. M. (2009). Agrupar Textos Cifrados é Equivalente a Agrupar Textos legíveis. VII Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana.

[22] Bible (2005), “Bible in basic english”, Disponível: http://www.o-bible.com/bbe.html [capturado 13 dez. 2005].