# Novas formas de ataques de distinção a cifradores: caminhos para o futuro

*Almir Paz de Lima*, José Antônio Moreira Xexéo** e Paulo Roberto Gomes****

***RESUMO***

*A identificação do algoritmo criptográfico partindo do conhecimento, unicamente, da cifra por ele gerada é uma aspiração de criptanalistas e uma preocupação que atormenta os projetistas desse tipo de algoritmo, destinados a assegurar o sigilo das informações.*

*Embora a identificação do algoritmo cifrador, por si só, não venha a revelar a informação embutida na cifra, ou a chave particular usada na transformação, ela evidencia a existência de uma "assinatura", isto é, um padrão que caracteriza o algoritmo usado. E o domínio da elaboração desse padrão poderá acarretar, mais cedo ou mais tarde, o desvendamento de fraquezas e a possível quebra do algoritmo. O avanço das pesquisas em reconhecimento de padrões, largamente usado para classificação e recuperação de informações, fornece um novo caminho para a busca de padrões em criptogramas, permitindo correlacioná-los com os possíveis parâmetros que os originaram, tais como o algoritmo criptográfico ou, até mesmo, a particular chave usada no processo. Os promissores resultados, preliminarmente obtidos, e a ausência de relatos na literatura especializada em criptologia permitem acreditar nesses novos caminhos para o futuro da criptanálise: o emprego de ferramentas de inteligência artificial, com apurada sensibilidade para pesquisar e identificar padrões complexos.*

## Introdução

A confiabilidade de algoritmos criptográficos, destinados que são a garantir o sigilo, a integridade e a autenticação da origem, é fundamental no moderno ambiente de comunicações globalizadas. A segurança das informações é fator imprescindível não somente para as comunica-

* Tenente-Coronel Ref Ex, Eng/ AMAN-56, Eng Com IIVIE-63, MC ITA-69, MSc Illinois-71, é pioneiro do ensino acadêmico de Criptografia no país, criou as Linhas de Pesquisa em Criptografia na Pós-graduação do IME em 1978.

** Tenente-Coronel Ref Ex, Com/AMAN-64, Eng Com/ME-72, MC IME-83, DC UFRJ-O1. É professor assistente e coordenador do curso de Sistemas de Informação da Faculdade Salesiana Maria Auxiliadora, avaliador institucional e de curso de graduação do SINAES, professor do IME e coordenador científico do grupo de pesquisa em segurança da informação e o gerente técnico do Grupo de Segurança da Informação do PBCT do DCT / EB.

*** Coronel RI, Com /AMAN-70, Eng Com IME-77, MC IIV1E-82, CDEM/ECEME-92. Integrante do GSUPBCT, é professor da Disciplina Segurança da Informação para o Curso de Eng Computação/IME.

ções comerciais como para as militares. A obtenção e posse de conhecimentos que permitam reduzir essa confiabilidade podem representar um importante fator de desequilíbrio nas relações comerciais ou de poder, num sentido mais amplo.

O comportamento estatístico dos algoritmos criptográficos, cujo tamanho da chave seja compatível com os atuais recursos computacionais, é uma medida inicial de confiabilidade. Assim, a avaliação em relação aos critérios de aleatoriedade das sequências geradas por esses algoritmos pode indicar, *a priori*, se ele não é seguro. Caso seja aprovado neste quesito, e enquanto estiver em uso, fraquezas específicas são pesquisadas, para verificar possíveis falhas de projeto que permitam minimizar o esforço da quebra, sem necessidade de experimentar todas as chaves possíveis (quebra por exaustão).

O desenvolvimento dos sistemas de reconhecimento de padrões e suas mais recentes aplicações permitem acreditar no sucesso do emprego de técnicas de inteligência artificial, também no esforço de criptanálise.

A busca de padrões em criptogramas, identificando o algoritmo criptográfico que o produziu e até a separação de conjunto de criptogramas gerados pela mesma chave, em um algoritmo determinado, aponta um promissor caminho, não só para auxiliar a criptanálise como também uma valiosa ferramenta para o projeto de cifradores confiáveis.

Neste artigo será apresentada uma descrição dessas pesquisas, conduzidas pelo Grupo de Segurança da Informação do Instituto Militar de Engenharia (GSUIIVIE), registradas em duas Dissertações de Mestrado [1, 2] e um artigo na revista *Cryptologia* [3], procurando realçar um expressivo resultado obtido: a identificação do Padrão Americano de Criptografia de Dados (DES), usado por cerca de 30 anos e substituído em 2001, mediante a análise de criptogramas por ele gerados misturado em um conjunto de criptogramas de outros algoritmos.

## Desenvolvimento

Para a descrição das pesquisas conduzidas pelo GSI/IME será caracterizado o problema a ser tratado neste trabalho, a indicação da solução normalmente adotada, a descrição da nova solução proposta e os resultados obtidos.

### *Caracterização do problema*

A identificação de padrões em criptogramas pode ser uma ferramenta poderosa para o esforço de criptanálise, e a compreensão de como se formam esses padrões pode levar à quebra do algoritmo criptográfico. Por outro lado, o domínio do comportamento dos parâmetros que acarretam esses padrões pode melhorar o projeto do algoritmo, tornando-o resistente à criptanálise.

A constatação de que um cifrador imprime uma assinatura nas cifras por ele geradas pode ser indício de fraqueza e até levar ao desvendamento da informação nele ocultada. A história da criptologia demonstra que toda descoberta de fraqueza em um sistema criptográfico desenboca na sua quebra e no surgimento de novo sistema, para corrigir a debilidade evidenciada.

Destarte, o sistema de substituição monoalfabética levou pouco tempo para exibir sua fragilidade estrutural: os caracteres do texto cifrado carregavam as medidas probabilísticas dos caracteres do texto claro aos quais substituíam; o sistema polialfabético, que o sucedeu, espalhava a frequência de ocorrência do caractere do texto claro por diferentes caracteres do texto

cifrado. Isso o tornou operante por mais de 300 anos, mas não resistiu à constatação da formação de padrões esparsos, percebidos por diligentes observadores [4].

A era computacional trouxe algoritmos sofisticados, que abandonaram o tratamento sobre os caracteres do texto claro e passaram a manipular os bites componentes dos códigos de representação desses caracteres. Aparentemente, então, de nada valem as informações sobre a frequência de ocorrência dos caracteres na linguagem correspondente ao texto claro, pois a difusão e a confusão conjugam bites de caracteres diferentes.

A suposta ausência de correlação dos caracteres do texto claro com os da cifra gerada, por causa da ausência de padrões detectáveis, permite tratar os cifradores computacionais como geradores de sequências aleatórias. Daí a medida de aceitação desses cifradores estar fundamentada no grau de aleatoriedade das sequências que compõem o criptograma. O exemplo mais marcante desse tratamento é o conjunto de testes proposto pelo Nacional Institute of Standard and Technology (NIST), como avaliação preliminar de cifradores [5].

***Ataques de distinção a cifradores***

Os testes relacionados com aleatoriedade, disponibilizados pelo NIST, foram usados no recente processo de escolha do padrão avançado de criptografia de dados do governo americano [6]. Por se tratar de medidas de aleatoriedade das sequências geradas, esse fato reforçou a imagem dos cifradores como geradores de sequências aleatórias, induzindo processos estatísticos de análise dos algoritmos criptográficos.

Medidas estatísticas distinguem um gerador de sequências aleatórias, com uma probabilidade de distribuição $p_x$ desconhecida, de um gerador de sequências aleatórias, com distribuição uniforme $p_u$, mediante um teste denominado de qui-quadrado. Este teste foi aplicado pela primeira vez, segundo está registrado na literatura, por Vaudenay sobre o Data Encryption Standard (DES), sendo chamado de "ataque de distinção" [7], em 1996 e, portanto, anterior à publicação dos testes do NIST,.

Ueda e Terada, em 2007, no VII Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais [8], fazem referência ao trabalho de Vaudenay como aos quatro próximos artigos a seguir citados.

Em 2000, Knudsen e Meier [8] descrevem o ataque de distinção ao RC6, um dos cinco candidatos finalistas à escolha do padrão avançado de criptografia de dados (AES) do governo americano, e uma variante desse mesmo algoritmo.

Em 2003, Ryabko [8] discorre sobre aplicações de teste qui-quadrado em criptografia.

Em 2004, Takenada, Shimoyama e Koshiba [8] fazem uma análise teórica sobre o ataque qui-quadrado ao RC6.

Em 2005, Miyaji e Takano [8], na Conferência Australiana em Informação, Segurança e Privacidade, discorrem sobre o sucesso de um ataque usando testes qui-quadrado sobre o RC6.

***Um novo caminho se configura***

Na procura de fraquezas em algoritmo criptográfico projetado para uso institucional, o GSI/IME enveredou por um caminho até então desconhecido, ao menos não publicado na literatura especializada. A proximidade com grupo de pesquisa em Linguística proporcionou uma aplicação inovadora de técnicas de recuperação de informações em criptanálise [1].

Em 2006, Oliveira, Xexéo e Carvalho publicaram um primeiro artigo na revista *Cryptologia* [3],

descrevendo uma aplicação de técnicas de agrupamento e classificação de informações em criptologia.

Em 2007, no trabalho seguinte do grupo, Souza [2] ampliou os resultados obtidos em Carvalho, realizando exaustivos experimentos para otimizar os parâmetros envolvidos nas diversas medidas. Além disso, usou redes neurais para classificar criptogramas, segundo o algoritmo e as chaves empregadas.

Desse trabalho de Souza foi selecionado um experimento, adiante descrito, para consolidar conceitos deduzidos em experimentos anteriores, e no qual se demonstrou que a partir de criptogramas é possível identificar os algoritmos criptográficos que os produziram.

Para o experimento foram gerados 1.200 criptogramas obtidos da criptografia da Bíblia, em inglês [9], separada em 30 subtextos de exatamente 1.024 baites (1 KB), pelos cifradores: DES, com 10 chaves de 64 bites (56 bites úteis) fornecendo 300 criptogramas; e AES, com 10 chaves de 128 bites, fornecendo 300 criptogramas, 10 chaves de 192 bites, fornecendo 300 criptogramas e 10 chaves de 256 bites, fornecendo 300 criptogramas.

Usando ferramentas de Inteligência Artificial, esse conjunto foi reagrupado em subconjuntos procurando identificar o cifrador/chave que os geraram, conforme o item a seguir.

***Descrição do experimento escolhido***

O texto da Bíblia, em inglês, foi submetido a um separador de textos que, a partir do parágrafo inícial, separou 30 subtextos (numerados de 1 a 30) de tamanho igual a 1 KB (disponíveis em http://www.cos.ufi7.br/-william/TextosL egiveis.rar).

Esses 30 textos claros foram cifrados no modo Livro Eletrônico de Códigos (ECB) pelos algoritmos DES e AES (disponíveis em http://www.cos.ufrj.br/-william/ TextosLegiveis.rar). A razão de se optar pelo modo ECB está no propósito de estudar o comportamento do conjunto algoritmo-chave e não de seu modo de operação.

Foram geradas, aleatoriamente, 10 chaves de 64 bites para o DES, 10 chaves de 128 bites para o AES_128, 10 com 192 bites para o AES-192 e 10 de 256 bites para o AES 256.

Cada um dos 30 textos foi cifrado com cada uma das 10 chaves para os quatro cifradores, obtendo-se 1.200 criptogramas.

Cada criptograma foi tratado de forma a ser representado no modelo Espaço de Vetores [ 10, 11 ], compondo o *corpus criptus*, pronto para ser submetido ao processo de agregação que, se atuasse com perfeição, deveria gerar 40 grupos de 30 criptogramas, correspondendo aos quatro cifradores com 10 chaves cada.

O processo de agregação valeu-se da medida de similaridade conhecida como Ângulo do Cosseno e para critério de parada foi adotado o valor de similaridade a partir de 0,001. A técnica de agrupamento foi a Hierárquica Aglomerativa e o método o Single-link [11].

O experimento foi conduzido em quatro fases. Cada fase considerou o elemento léxico básico do criptograma, chamado "criptotermo" e equivalente à palavra do texto claro, como sendo de tamanho diferente. Na fase-1, o criptotermo foi de 64 bites, na fase-2, de 128, na fase-3, de 192 e, na fase-4, de 256, seguindo o tamanho das chaves requeridas pelos algoritmos usados.

***Resultados do experimento***

Na fase-1, com criptotermos de 64 bites, os resultados foram:

— para o DES foram separados 10 grupos com 30 criptogramas cada, correspondendo

cada grupo a uma das 10 chaves usadas. Tanto a medida de precisão quanto a taxa de recuperação foram de 100%, significando que todos os elementos foram corretamente classificados e todos os elementos que se desejava juntar foram parar no mesmo grupo, ou seja, sucesso absoluto;

— para o AES_128 foram obtidos 220 grupos, sendo 22 para cada chave usada e, desses 22, 18 grupos de um único criptograma ( correspondentes aos textos claros 3, 4, 5, 6, 7, 9, 10, 12, 13, 14, 15, 18, 20, 21, 22, 23, 27, 30), dois grupos com dois criptogramas cada (correspondentes aos pares de texto claro 1-2 e 11-29), um grupo com três criptogramas (correspondentes aos textox claros 16-17-19) e um grupo com cinco criptogramas (correspondentes aos textos claros 8-24-25-26-28). Essa associação de grupos se repetiu em todas as chaves, levando à conjectura, a ser explorada em próximas pesquisas, de que a afinidade pode ter sido influenciada, também, pelos textos claros que os originaram,

— tanto para o AES_192 quanto para o AES 256 o comportamento foi exatamente o mesmo do AES_128, reforçando a conjectura de que a chave não influiu na agregação.

Na fase-2, com criptotermos de 128 bites, os resultados foram:

— para o DES e os três AES foram separados 22 grupos para cada chave com o padrão exatamente igual ao descrito para os três AES da fase-1. É interessante notar que foi perdida a sensibilidade, para o DES, demonstrada com o criptotermo igual a 64 bites da fase-1. Por outro lado, não aumentou a sensibilidade, que a intuição poderia apontar, para o AES_128, mostrando que o tamanho do criptotermo é um fator crítico no processo de agregação. Pode ser admitido, ainda, que o critério de parada com valor de similaridade a partir de 0,001 deva ser reduzido em função do tamanho do criptotermo, além de aumentar o tamanho do texto a ser cifrado.

Tanto na fase-3 como na fase-4 foram obtidos 1.200 grupos, não havendo nenhuma agregação, nem mesmo as detectadas com os parâmetros anteriores, e que as pesquisas necessitam ser aprofundadas.

**Conclusão**

Comparado aos testes estatísticos atualmente empregados, como o do qui-quadrado, que aceita o cifrador como gerador de sequências aleatórias, dentro de intervalos de confiança previamente estabelecidos, o método descrito neste trabalho representa um novo paradigma, que pode levar a resultados difíceis de se prever nestes primeiros passos da pesquisa.

Na fase-1 do experimento está sintetizada a força do método. Embora o objeto de estudo tenha sido o DES, já substituído como padrão de criptografia pelo Rijndael [12], deve-se lembrar que permaneceu em uso por cerca de 30 anos, sofrendo os mais variados tipos de ataques criptanalíticos. E o sucesso do experimento foi de 100%, incluindo o agrupamento correto dos criptogramas obtidos por chaves diferentes. Este fato é de grande relevância, pois relaciona chaves usadas com os padrões detectados.

Das demais fases pode-se concluir que os parâmetros envolvidos, tais como tamanho do texto, criptotermo, critério de parada, correlação dos textos claros, dentre outros, precisam ser mais bem manipulados, para se chegar a um conhecimento mais profundo sobre seus comportamentos.

As pesquisas prosseguem, visando a um refinamento dos resultados. Ao se ajustar os parâmetros em função de algoritmo-chave poder-se-á

adotar uma metodologia similar ao processo de garimpagem, com a aplicação de peneiras cada vez mais finas:

— em uma primeira etapa, retirar os criptogramas gerados pelo DES, se houver;

— reagrupar os criptogramas e reestruturar o *corpus criptus*;

— repetir a fase anterior, agora para o AES-128, ou o próximo algoritmo do grupo, até que todos sejam identificados.

## Referências


[1] CARVALHO C. A. B. de (2006). O Uso de Técnicas de Recuperação de Informações em Criptoanálise. Dissertação de Mestrado – Instituto Militar de Engenharia.

[2] 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/tesel ME.pdf

[3] OLIVEIRA C. M., XEXÉO J. A. M. e CARVALHO C. A. B. (2006) Clustering and Categorization Applied to Cryptology. Cryptologia vo130, p 266-280, 2006.

[4] SING, S. (2001), O Livro dos Códigos; A ciência do sigilo – do antigo Egito à criptografia quântica. Editora Record.

[5] RUKHIN A. et al.(2001), "A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications", NIST.

[6] SOTO (1999). "Randomness Testing of the Advanced Encryption Standard Candidate Algorithms", NIST.

[7] VAUDENAY S. (1996) An Experiment on DES Statistical Cryptanalysis. ACM Conferenceof Computer and Communications Security, pages 139-147.

[8] UEDA, T. K. e TERADA, R. (2007). Uma Versão Mais Forte do Algoritmo RC6 contra a criptoanálise $x^2$. VII Simpósio Brasileiro em Segurança da Informação e de Sisteams Computacionais.

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

[10] RASMUSSEN, E. (1992), "Clustering algorithms". In Information retrieval: data structures and algorithms, Edited by William Frakes and Ricardo Yates, Prentice Hall, p. 419-442.

[11] YATES, R.B. e NETO, B. R. (1999), Modern information retrieval. Addison

[12] NIST (2001). Federal Information Processing Standard, publication 197 (FIPS 197): Announcing the advanced encryption standard (AES). Washington D. C.