

Trabalho 1 de Segurança de Sistemas

1st Arthur Sudbrack Ibarra

Pontifícia Universidade Católica do Rio Grande do Sul
Porto Alegre, Rio Grande do Sul, Brasil
arthur.ibarra@edu.pucrs.br

2nd Giovanni Gonçalves Migon

Pontifícia Universidade Católica do Rio Grande do Sul
Porto Alegre, Rio Grande do Sul, Brasil
giovanni.migon@edu.pucrs.br

Resumo—Este trabalho propõe uma solução para decifrar textos criptografados com a cifra de Vigenère, como parte do primeiro trabalho da disciplina de Segurança de Sistemas. O algoritmo desenvolvido analisa exclusivamente o texto cifrado para determinar tanto o tamanho da chave quanto a chave real utilizada na criptografia. Utilizando índices de coincidência e análise de frequência de letras, o algoritmo oferece uma possível abordagem para decifrar textos cifrados com essa técnica.

Palavras-chave—Segurança, Cifra, Vigenère, Criptografia, Decodificação

I. VÍDEO DE DEMONSTRAÇÃO DO ALGORITMO

https://youtu.be/eMxpPk7n_LA

II. INTRODUÇÃO E PROBLEMA

A cifra de Vigenère [1] é um método clássico de criptografia que utiliza uma chave para cifrar textos. Neste trabalho, é proposto um algoritmo para decifrar textos criptografados com a cifra de Vigenère sem conhecimento prévio da chave. A cifra de Vigenère funciona cifrando o texto original por meio de deslocamentos de letras, semelhante à cifra de César [2]. No entanto, o diferencial da cifra de Vigenère está na utilização de uma chave de tamanho variável, que se repete para cifrar o texto. Dessa forma, diferentes partes do texto são cifradas utilizando letras diferentes da chave, aumentando a complexidade da criptografia.

O algoritmo proposto neste trabalho emprega uma abordagem baseada em índices de coincidência e análise das frequências das letras para determinar o tamanho da chave e, subsequentemente, a chave real utilizada na criptografia. Ao analisar as características estatísticas do texto cifrado, o algoritmo é capaz de identificar padrões que auxiliam na determinação do tamanho da chave e na inferência da língua em que o texto foi escrito.

III. SOLUÇÃO

Nesta seção, será explicado o funcionamento do algoritmo proposto para decifrar textos criptografados com a cifra de Vigenère. A explicação a seguir abordará tanto a determinação do tamanho da chave quanto a descoberta da chave real utilizada na criptografia.

A. Determinação do Tamanho da Chave

Para decifrar textos criptografados com a cifra de Vigenère, o algoritmo desenvolvido emprega uma abordagem sistemática baseada em índices de coincidência e análise das frequências das letras. Inicialmente, o texto cifrado é dividido em vários

subtextos, cada um correspondendo a uma suposição diferente para o tamanho da chave. Essa divisão é fundamental porque, dessa forma, podemos garantir que cada letra de um subtexto foi cifrada pela mesma letra da chave, uma vez que a chave se repete.

Para cada subtexto, são calculados os índices de coincidência, que representam a probabilidade de duas letras escolhidas aleatoriamente no texto serem iguais. Esses índices são então comparados com os valores típicos para os idiomas inglês e português, que são de aproximadamente 0.066 e 0.074, respectivamente [3]. Além disso, a frequência das duas letras mais comuns em cada subtexto é analisada para verificar se correspondem às frequências típicas dos idiomas. Essas verificações não apenas permitem estimar o tamanho da chave, mas também identificar se o texto criptografado era originalmente escrito em português ou em inglês.

B. Descoberta da Chave

Após determinar o tamanho da chave, é necessário descobrir a chave em si. Para isso, cada subtexto é analisado individualmente para identificar a letra mais frequente. Com base na língua identificada (inglês ou português), as duas letras mais frequentes no idioma são consideradas para realizar os deslocamentos necessários. Nesse contexto, o deslocamento refere-se à aplicação da cifra de César, que consiste em deslocar cada letra do texto cifrado para trás no alfabeto um número fixo de posições, determinado pela chave.

A exemplo do que foi mencionado, em um dos textos cifrados fornecidos pelo professor, foram identificadas as possíveis chaves “hezmh” e “david”. A chave “hezmh” foi assumida considerando que as letras mais frequentes em todos os subtextos representavam o caractere “A” em português, que é a letra mais comum. Já a chave “david” foi assumida considerando que as letras mais frequentes representavam o caractere “E”, que é a segunda letra mais comum.

Com base nas chaves identificadas, é possível gerar todas as combinações possíveis, incluindo as permutações dessas chaves principais. A Figura 1 ilustra visualmente esse processo de permutação e as chaves resultantes. Todas essas combinações são consideradas como possíveis chaves para decifrar o texto cifrado, permitindo explorar várias hipóteses e aumentando a probabilidade de encontrar a chave correta. Embora a chave correta possa não estar entre as geradas, essa ocorrência é extremamente rara, pois nem a primeira nem a segunda letra

The diagram shows two keys at the top: **Chave 1** (hezmh) and **Chave 2** (david). A vertical line connects them, with an arrow pointing down to a box containing a list of other possible keys:

- hezih, hezid, hevmh, hevmd, hevih, hevid,
- hazmh, hazmd, hazih, hazid, havmh, havmd,
- havih, havid, dezmh, dezmd, dezih, dezeit,
- devmh, devmd, devih, devid, dazmh, dazmd,
- dazih, dazid, davmh, davmd, davih, hezmd

Outras Possíveis Chaves

É importante ressaltar que no exemplo mencionado, a chave verdadeira era *david*, representando um caso incomum em que todas as letras mais frequentes dos subtextos se mapearam para o “E”. Normalmente, espera-se que a letra mais comum seja o “A” em pelo menos algum dos subtextos. No entanto, neste caso específico, a frequência do “E” foi maior em todos os subtextos. As frequências típicas do “E” e “A” no alfabeto português são 12.570% e 14.634%, respectivamente [4].

A eficácia do algoritmo desenvolvido foi demonstrada através da sua aplicação a uma variedade de textos cifrados. Os resultados ilustram notavelmente a capacidade do algoritmo em adaptar-se a diferentes idiomas e contextos. Na Figura 2, observamos a decifração bem-sucedida de um texto em inglês. A chave identificada, “meunome“, de tamanho 7, foi descoberta ao mapear as letras mais frequentes de cada subtexto para o “E“, a letra mais comum no alfabeto inglês. O processo revelou um *e-book* completo em inglês.

Já a Figura 3 exibe a decifração de um texto cifrado em português, onde a chave encontrada foi “david“, de tamanho 5. Este resultado foi alcançado ao considerar que as letras mais frequentes em todos os subtextos correspondiam ao “E“ do alfabeto português, conforme comentado na seção III-B, levando à revelação de um trecho da Bíblia Sagrada.

Por fim, a Figura 4 apresenta um caso intrigante onde o texto cifrado em português exibiu um padrão de frequência de letras variado. Alguns subtextos apresentaram o “A“ como letra mais frequente, enquanto outros indicaram o “E“. Este cenário exigiu uma abordagem mais refinada, levando à descoberta da chave “meunome“ novamente, de tamanho 7, após extensas permutações discutidas na Figura 1. O texto decifrado mostrou-se ser parte de um livro em português.

Figura 2. Decifragem de Texto Cifrado em Inglês

Figura 3. Decifragem de Texto Cifrado em Português - 1

Figura 4. Decifragem de Texto Cifrado em Português - 2

O algoritmo desenvolvido mostrou-se capaz de decifrar textos cifrados com a cifra de Vigenère de forma satisfatória. Ao analisar as características estatísticas do texto cifrado, o algoritmo conseguiu determinar o tamanho da chave e identificar a chave real utilizada na criptografia. Embora casos excepcionais possam ocorrer, nos quais a chave correta não é identificada, essas situações são raras e geralmente estão associadas a particularidades na distribuição das letras no texto cifrado.

Esta análise ressalta, ainda, a fragilidade da cifra de Vigenère frente aos métodos modernos de criptoanálise. Ao explorar suas vulnerabilidades através de técnicas estatísticas e de análise de padrões, demonstrou-se que a segurança dessa cifra é facilmente comprometida. A repetição periódica da chave e a falta de difusão adequada resultam em padrões previsíveis no texto cifrado, tornando-o suscetível a ataques de decifração.

[1] Cifra de César. (2024). Acesso em 24 de Março de 2024, disponível em https://pt.wikipedia.org/wiki/Cifra_de_C%3%A9sar

- [2] Cifra de Vigenère. (2024). Acesso em 24 de Março de 2024, disponível em https://pt.wikipedia.org/wiki/Cifra_de_Vigen%C3%A8re
- [3] Index of coincidence. (2024). Acesso em 24 de Março de 2024, disponível em https://en.wikipedia.org/wiki/Index_of_coincidence
- [4] Letter frequency. (2024). Acesso em 25 de Março de 2024, disponível em https://en.wikipedia.org/wiki/Letter_frequency