

Uma Análise Comparativa dos Padrões entre Senhas Brasileiras e de Língua Inglesa

Marino Souza S., Nilton V. C. Júnior, Luiz R. Rios

¹Instituto de Matemática – Universidade Federal da Bahia (UFBA)

{marino, niltonvasques}@openmailbox.org, luizromario@gmail.com

Resumo. *O método de adivinhação de senhas por força bruta necessita de um bom dicionário a fim de conseguir o maior número de advinhações possíveis. Motivando-se no trabalho de [Li and Han 2014]. O presente artigo visa executar uma análise em cima de senhas web com o objetivo de encontrar padrões que futuramente sirvam para melhorar dicionário de adivinhação, ou desenvolver melhores políticas de senhas.*

1. Introdução

As bases de senhas utilizadas neste trabalho foram encontradas em sua maior parte no *website SkullSecurity* (<https://blog.skullsecurity.org/>) contendo senhas de provável usuários de língua inglesa, e outra base disponibilizada também de forma pública contendo senhas de um órgão de defesa de um governo de um país que neste trabalho terá o nome omitido, utilizando o codinome: BrArmy. A quantidade de senhas totalizaram 64,493 providas do *SkullSecurity* e 7,834 do BrArmy. Em ambas as bases de senhas haviam senhas em branco e com charset inválido que podem afetar as análises. Após a filtragem de alguns desses dados a quantidade total caiu para 64,463 da *SkullSecurity* e 7,833 da BrArmy.

Durante este trabalho os autores preferiam utilizar o sistema internacional para representação de números reais, separando por vírgula, e múltiplos de mil, separando por ponto.

2. Estatísticas Comuns

fazer uma explicação da sessão.

	SkullSecurity	BrArmy
1	123456(0.202%)	12345678(4.864%)
2	password1(0.119%)	123456789(1.009%)
3	fuck(0.092%)	87654321(0.230%)
4	abc123(0.090%)	10203040(0.204%)
5	fuckyou(0.064%)	06121966(0.153%)

Table 1. Senhas mais usadas nas duas bases

Explicar oq a tabela 1 significa.

Curiosamente, entre as senhas de língua inglesa é mais comum termos palavras escatológicas, como apresentado em 2, enquanto os usuário da base BrArmy recorrem, em sua maioria, a palavras que tem alguma relação as forças armadas.

	SkullSecurity	BrArmy
1	password1(0.119%)	flamengo(0.115%)
2	fuck(0.092%)	exercito(0.115%)
3	fuckyou(0.064%)	infantaria(0.115%)
4	monkey(0.045%)	cavalaria(0.102%)
5	iloveyou1(0.043%)	guilherme(0.077%)

Table 2. Palavras Inglesas/Portuguesas mais usadas

2.1. Composição e Estruturas das senhas

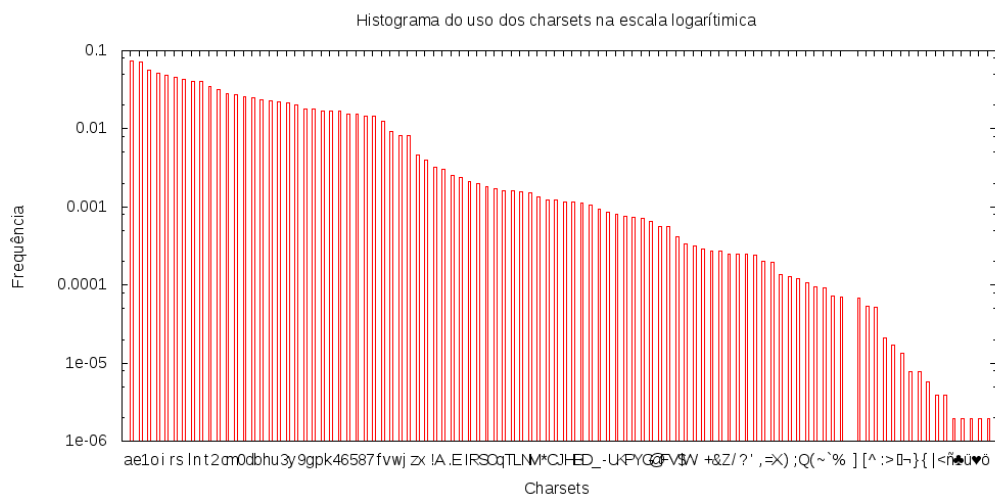


Figure 1. Senhas da base SkullSecurity.

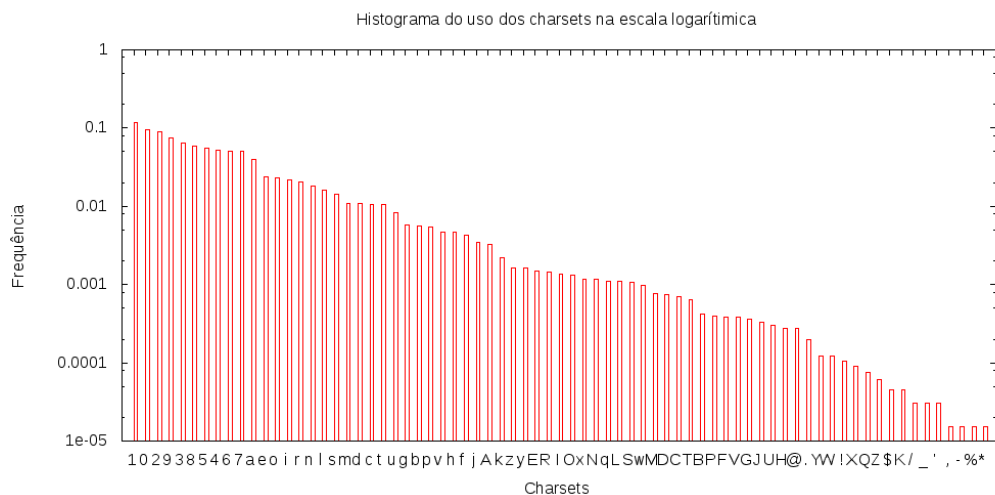


Figure 2. Senhas da base BrArmy

Nas imagens 1 e 2 temos a distribuição de cada caractere em cada base. É fácil notar que os valores mais altos da 2 são responsáveis pelos caracteres numéricos, enquanto na 1 temos uma maior distribuição destes.

	digit	lowercase	lowercase +digit	lowercase +symbol	digit +symbol	lowercase +digit+symbol	samerow
SkullSecurity	6.15%	23.89%	56.93%	4.73%	0.08%	2.04%	0.64%
BrArmy	61.13%	11.44%	23.22%	0.09%	0.06%	0.26%	5.91%

Table 3. Composição das senhas.

	Estrutura Mais Comum	%	Segunda Mais Comum	%
SkullSecurity	LLLLLL	6.874%	LLLLLLD	6.208%
BrArmy	DDDDDDDD	51.845%	DDDDDDDDDD	6.485%

Table 4. Estruturas Mais Comuns da senhas.

motivar os proximos dados

Na tabela 4 temos as estruturas escritas com L e D, que representam letra e número, respectivamente. Abaixo temos a porcentagem de cada uma.

motivar as tabelas de datas

	Exatamente Oito Dígitos	DDMMYYYY	MMDDYYYY	YYYYMMDD
SkullSecurity	638(0.990%)	25.547%	5.799%	2.978%
BrArmy	3,565(45.513%)	26.928%	10.659%	0.701%

Table 5. texto da tabela

	Exatamente Seis Dígitos	DDMMYY	MMDDYY	YYMMDD
SkullSecurity	1,066(1.654%)	36.210%	19.325%	11.445%
BrArmy	0%			

Table 6. texto da tabela

A comparação entre os dados apresentados nas tabelas 5 e 6 são muito distintas, enquanto na primeira vemos uma quantidade enorme de datas provindas da base BrArmy, na segunda não é encontrado nenhum padrão de data de exatamente seis dígitos (DDDDDD), embora haja apenas uma senha no formato DD.DD.DD e mais uma no formato DD/DD/DD.

3. Considerações Finais

4. References

References

Li, Z. and Han, W. (2014). A large-scale empirical analysis of chinese web passwords. *23rd USENIX Security Symposium*, pages 558–574.