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

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

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

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

Resumo. *O método de adivinhação de senhas por força bruta pode alcançar maior eficiência se faz uso de um dicionário de ataque com padões de senhas. Motivando-se no trabalho de [Li and Han 2014] este trabalho visa executar uma análise em cima de senhas web com o objetivo de encontrar padrões que futuramente sirvam para melhorar dicionários 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* [SkullSecurity 2014] contendo senhas de provavéis usuários de lingua inglesa, e outra base dispinibilizada também de forma pública contendo senhas de um orgão de defesa de um governo de um país que neste trabalho terá o nome omitido, utilizando o codnome: `BrArmy`. A quantidade de senhas totalizaram 64,493 provindas do *SkullSecurity* e 7,834 do `BrArmy`. Em ambos 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 multiplos de mil, separando por `ponto`.

2. Estatísticas Comuns

Entre as mais de 70 mil senhas analisadas, uma diferença enorme se destaca na composição destas. As tabelas abaixo mostram algumas dessas diferenças.

	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

Na tabela 1 é mostrado as senhas mais usadas e o seu respectivo percentual, na primeira colona há a disposição das cinco senhas mais usadas entre as bases do `SkullSecurity` e na segunda é disposto as senhas da base `BrArmy`. Enquanto as

	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

senhas de provaveis usuários de língua inglesa são mais mistas (letras e números) as senhas de usuários de língua portuguesa usam mais da estrutura unicamente numérica.

Curiosamente, entre as senhas de língua inglesa é comum encontrarmos 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.

2.1. Composição e Estruturas das senhas

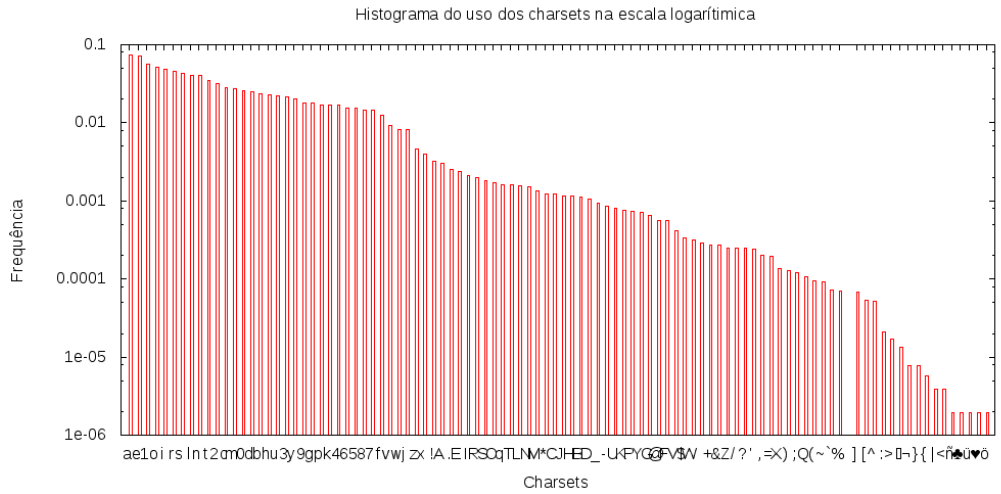

Figure 1. Senhas da base SkullSecurity.

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 caracters numéricos, enquanto na 1 temos uma maior distribuição destes.

	digit	lowercase	lowercase +digit	lowercase +symbol	digit +symbol	lowercase +digt+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.

A tabela **??** apresenta quais as principais composições das senhas das duas bases tratadas. A distribuição das colunas indica as categorias numérica, letra minúscula, e mistas com e sem caracteres especiais. A última coluna indica o percentual de senhas com caracteres na mesma linha.

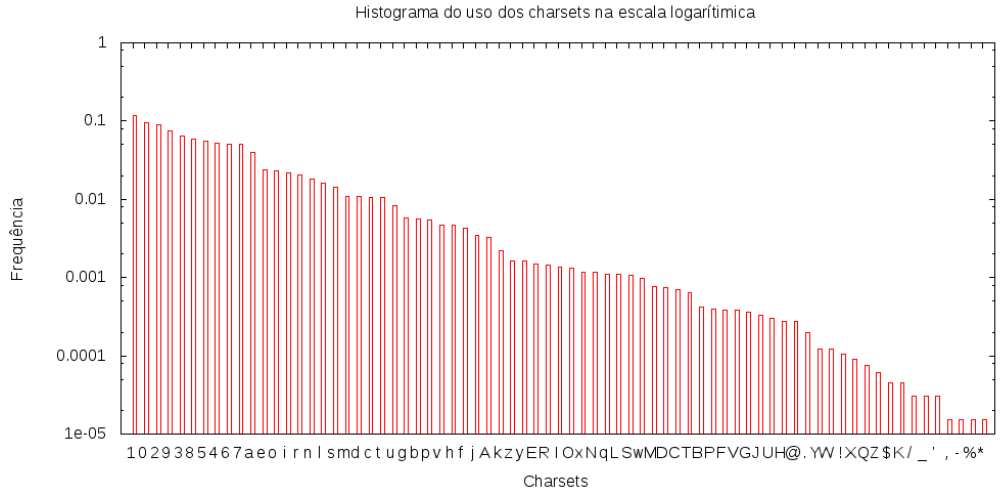

Figure 2. Senhas da base BrArmy

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

Table 4. Estruturas Mais Comuns da senhas.

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

Após essa exposição de grande percentual de senhas numéricas, vale saber qual o padrão que estas senhas seguem. Nos dados a seguir são analisados apenas as senhas que são suspeitas de seram 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. Proporção das senhas de oito dígitos.

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

Table 6. Proporção das senhas de seis dígitos.

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

Tais dados apresentados neste trabalho não são suficientes para concluir quais usuários daquele idioma possuiem senhas mais facéis de adivinhação, porém é um dado que deve ser considerado no momento que se pretende criar uma política de senhas mais fortes, não permitindo que os usuários definam senhas dos padrões mais comuns. Por outro lado, estes dados também servem para fortalecer um dicionário de ataque, a fim de aumentar a eficiêcia de uma adivinhação por força bruta.

Todas as etapas deste trabalho encontram-se disponíveis em [Souza et al. 2015].

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.

SkullSecurity (2014). Skull security download leaks. `https://blog.skullsecurity.org/`.

Souza, M., Vasques, N., and Rios, R. (2015). Repositorio deste papper. `https://github.com/Marinofull/analysis-password-leak`.