

# Password Guessing usando Machine Learning

En un mundo profundamente interconectado como el moderno, la cuestión de la protección de la identidad digital tiene una gran importancia. Por esta razón, el desafío de crear contraseñas confiables es un reto que se ha hecho cada vez más complicado.

## Redes Neuronales

Password	Probability
oblivion	0.002909
asdf1234	0.002900
iloveyou	0.001844
abcd1234	0.001365
hello123	0.001359
internet	0.001254
adamadam	0.001227
password	0.001176
warcraft	0.000705
asshole1	0.000549

La idea general detrás del uso de Redes Neuronales para descifrado de contraseñas se basa en modelos para predecir un carácter siguiente a partir de caracteres anteriores. La arquitectura del modelo inicia con una capa convolucional, seguida de esta capa, encontramos una capa de dropout, luego una capa densa seguida por otra capa convolucional. Luego realiza global average pooling y termina con una capa densa. El modelo es capaz de generar alrededor de 100 contraseñas por segundo.

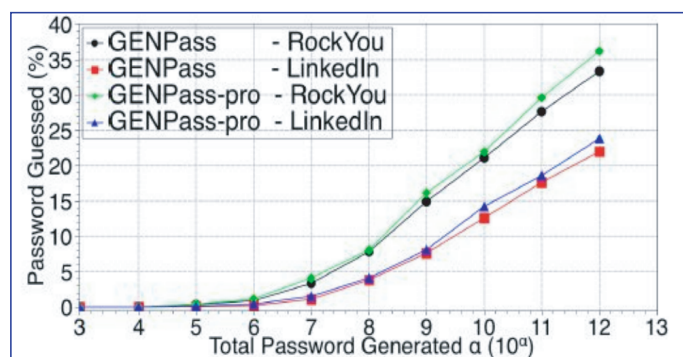
## PassGAN

Es un proyecto que usa deep learning y redes GAN en tensorflow para generar contraseñas. La red fue entrenada con 2 datasets: Rockyou ( 32'503,388) y LinkedIn (60'065,486). PassGAN es capaz de aprender los comportamientos que presentan las contraseñas que se basan en heurísticas.

Approach	(1) Unique Passwords	(2) Matches	(3) Number of passwords required for PassGAN to outperform (2)	(4) PassGAN Matches
JTR Spyderlab	$10^9$	461,395 (23.32%)	$1.4 \cdot 10^9$	461,398 (23.32%)
Markov Model 3-gram	$4.9 \cdot 10^8$	532,961 (26.93%)	$2.47 \cdot 10^9$	532,962 (26.93%)
HashCat gen2	$10^9$	597,899 (30.22%)	$4.8 \cdot 10^9$	625,245 (31.60%)
HashCat Best64	$3.6 \cdot 10^8$	630,068 (31.84%)	$5.06 \cdot 10^9$	630,335 (31.86%)
PCFG	$10^9$	486,416 (24.59%)	$2.1 \cdot 10^9$	511,453 (25.85%)
FLA $p = 10^{-10}$	$7.4 \cdot 10^8$	652,585 (32.99%)	$6 \cdot 10^9$	653,978 (33.06%)

## GENPass

Es una propuesta que da solución a unos problemas de PassGAN: Generalidad y exactitud. La exactitud se soluciona mediante un modelo PCFG+LSTM. La generalidad se soluciona con modelo basado en deeplearning entrenado con varios datasets al tiempo.



## Autor(es)

Daniel Amaris- ddamarisr@unal.edu.co  
Michael Guerrero- meguerreroa@unal.edu.co  
Christian Ortiz- cfortizp@unal.edu.co  
Jorge Camargo- jecamargom@unal.edu.co

## Organiza

Universidad Nacional de Colombia  
Grupo de investigación UNsecureLab  
Semillero de investigación Uqbar  
Grupo de investigación TLón