Sidharth Bambah

CS 35L

TA: Diyu Zhou

UID: 904 787 435

6 December 2017

Artificial Intelligence and Password Cracking

In modern computing, it has become vital to protect sensitive information. Methods of protection are numerous and vary widely, a common example being the use of public and private keys to encrypt information. However, a large majority of digital data is secured with a user provided password. With passwords comprising the majority of encryption for sensitive materials, there has been a rise in password cracking attempts. In the past, this hacking has been driven through programs, such as hashCat, which are built through complicated algorithms. Now, the use of artificial intelligence in the form of neural networks has been used to create a new software, PassGan, that can crack passwords much more easily.

Over the past few decades, password cracking has become ever more rampant. While numerous pieces of software and algorithms have been developed to breach passwords, one comes out on top: hashCat. This software was developed by Jens Steube and claims to be the world's fastest password recovery tool. The development of hashCat required many years to build the codebase and develop attack methods to successfully crack passwords. In order to run properly, hashCat takes a choice of algorithm that implement one of the three following methods: brute force, extrapolating, and probability. In the brute force method, the software attempts to “randomly try lots of combinations of characters” until the right password is found (Hutson). The extrapolation method guesses the correct password based on previously leaked ones. Finally, the probability method guesses each character of a password based on what came before. Along with a choice of algorithm, the hashCat software must be provided rules for the password cracking attempts. These rules are stored in a “wordslist” file and can be used to define schemes including, but not limited to, capitalization, special characters, and word combinations. Although the hashCat software has been quite successful, it does take a long time for the software to crack passwords. Thus, alternative development has begun to create software the can much more quickly, and accurately, crack passwords.

This new development has resulted in PassGan, a software that uses neural networks to crack passwords. PassGan is developed by a team of researchers at the Stevens Institute of Technology and generates numerous password guesses based on an input set. Some key differences of this software to hashCat is that PassGan does not require any manual analysis, meaning no long, complicated algorithms, and employs artificial intelligence to emulate how humans think (Hitaj). This software has been inspired by a subset of artificial intelligence, called deep learning. Deep learning involves neural networks and their development in layers. These neural networks learn in an unsupervised manner and are trained on patterns. Thus, they become much more successful and accurate when large amounts of data need to be analyzed. Furthermore, these networks are based on the human brain’s use of layers of neurons and they “learn” on training sets of data. These networks essentially create their own rules reducing the need for the user to enter his/her own rules into the software. Neural networks are often used in a “black box manner” according to Wojciech Samech from the Fraunhofer Institute of Telecommunications in Berlin (Snow). A specific subset of neural networks used in the PassGan software is the generative adversarial network, or GAN. This was introduced by Ian Goodfellow and consists of two, contesting neural networks: discriminative and generative. As can be seen by Figure 1, this network consists of the two neural nets mentioned previously.

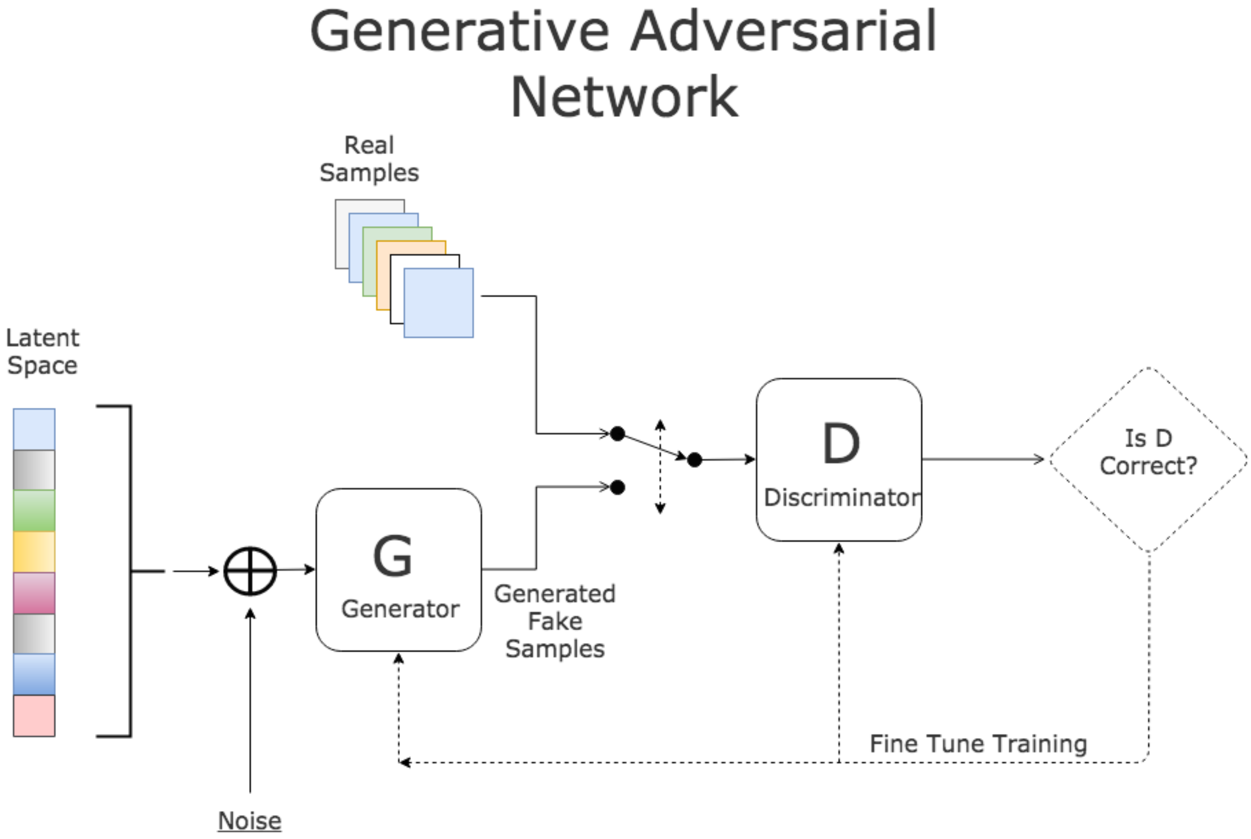

Figure 1. Generative Adversarial Neural Network. System consists of a generative and discriminative neural network. The real data and generated data are fed into the discriminator, which determines if they are a match.

The generative net takes in some inputs and creates an output that is passed into the discriminator, along with some true data. Then, the discriminator checks if the two inputs are match and returns feedback to the generator. In this way, the generative neural network "learns" based on the real inputs and sort of creates its own rules. This system is employed in PassGan, which takes in a set of leaked passwords and generates hundreds of millions of potential passwords. These new passwords are sent to a discriminative network and compared to the real passwords. This has proved to be quite successful and the generated passwords only get better as more and more data is collected. As mentioned, these neural networks “learn” with time and test data iterations. The researchers developing this software also tested PassGan to illuminate its potential. First, they fed the software the tens of millions of leaked passwords from the gaming website RockYou. Then the software created many new passwords that were compared to a list of leaked passwords from LinkedIn. The software generated 12% matches with the LinkedIn passwords suggesting that the neural network approach is quite successful in generating potential passwords (Hutson). Clearly, PassGan’s future as the next password cracking software is looking bright. However, it is important to take into account the two possible uses of such a software. Obviously, it can be used for nefarious purposes, but, as best stated by Thomas Ristenpart from Cornell, the software can be used for good by “generat[ing] decoy passwords to detect breaches”. Evidently, PassGan has quite a bright future and, with sustained testing, will be able to surpass hashCat and reduce the need for complicated algorithms and excessively powerful computers in the field of password cracking.

With the increase in software used to infiltrate passwords, there are still a few ways to keep crucial data safe. Some primary ways involve following the conventional wisdom of choosing passcodes that are 8 characters or more. Additionally, it would be wise to utilize a password that is a combination of letters, numbers, and even symbols. Along with these guidelines for strong passwords, multi-factor authentication is a good way to keep sensitive data secure. By paring accounts and passwords with smartphones, watches, and other devices, locks to digital files will only be able to be opened when the user has the correct password and the ability to access the paired device(s). Finally, it may be possible to use other methods of encryption along with passwords to keep data safe. For example, a public and private key method can be employed along with a complex password to allow files to be tagged with a proper owner. Despite the myriad ways to protect files, artificial intelligence has the potential to do great, and potentially nefarious, things.

Works Cited

Hutson, Matthew. “Artificial Intelligence Just Made Guessing Your Password a Whole Lot Easier.” ScienceMag, 15 Sept. 2017, www.sciencemag.org/news/2017/09/artificial-intelligence-just-made-guessing-your-password-whole-lot-easier

Hitaj, Briland, et al. “PassGAN: A Deep Learning Approach for Password Guessing.”*Cornell Univsersity Library*, 1 Sept. 2017, www.arxiv.org/abs/1709.00440

Snow, Jackie. “Brainlike Computers Are a Black Box. Scientists Are Finally Peering Inside.” ScienceMag, 26 July 2017, www.sciencemag.org/news/2017/03/brainlike-computers-are-black-box-scientists-are-finally-peering-inside