For today's learning, I will be researching on the contents of the paper "SoK: Machine vs. machine – A systematic classification of automated machine learning-based CAPTCHA solvers" that was published this month in the "Computers & Security" journal.

CAPTCHA's contributions to security

With automated "bots" instructed to do all kinds of unwanted activities, web applications have turned to a type of controls named "CAPTCHA" to stop these activities. CAPTCHA controls are designed to stop these "bots" since these "bots" are actually software programs on the internet.

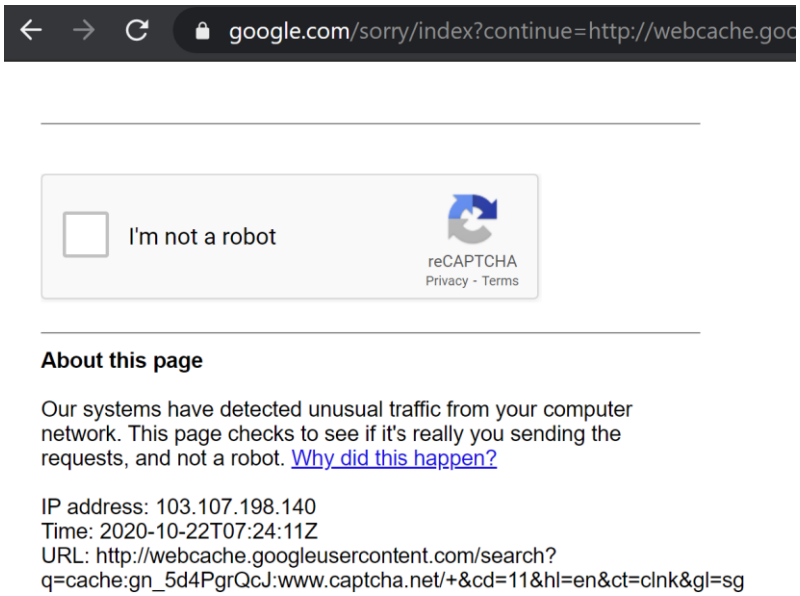

Figure: An example of a captcha – Google' reCAPTCHA

With relation to infosecurity, CAPTCHAs are generally used to prevent automated submissions (while allowing legitimate human submissions). Some of the submissions a CAPTCHA can be used to reduce or prevent are:

1. Password guessing using a dictionary/brute force
2. Automated searches in order to scrape prices and other product details
3. False comments or posting on forums or social media
4. Fake registration of services
5. Malicious poll responses (making it seem that a large portion of the public supports a particular stand)
6. Falsifying torrent seed counts so that more victims will download a particular malware

Several implementations exist, including:

- Text-based CAPTCHAs
- "Invisible" CAPTCHAs (e.g. Google reCAPTCHA)
- Maths captchas
- Audio captchas

Defeating CAPTCHA using Machine Learning

With most CATPCHAs being text-based, we will look into the general steps for a anti-CAPTCHA software to bypass the CAPTCHA software. For purpose of this article, I will only be discussing a direct attack on the generated image and not other weaknesses in implementation. I will leave it to the reader to imagine the scenarios where no direct attack is needed – one example, a CAPTCHA whose answers are general knowledge (obtainable from a database) or with little possibilities (such a birth years).

Machine Learning is generally thought of to be effective in bypassing current CAPTCHA implementations. A review was done on 51 papers and the following conclusions:

1. Machine Learning have been and will increase in accuracy, speed and abstraction in getting a solution to solve CAPTCHAs.
2. Reinforcement Learning, Generative Adversarial Networks (GANs), and Recursive Cortical Networks (RCNs) are better algorithms in solving CAPTCHAs. Other algorithms that are not based on Artificial Neural Networks(ANNs) are not as good as those.
3. A mix of algorithms will lead to maximum effectiveness in defeating CAPTCHAs. On the other hand, A mix of defensive techniques will also lead to optimal CAPTCHA effectiveness.

Definitions

I realize some of you may be a bit lost here, so here is a very short summary of the different terms (simple English from Wikipedia, where available):

1. Reinforcement learning (RL) is teaching a software agent how to behave in an environment by telling it how good it's doing. Reinforcement learning is different from supervised learning because the correct inputs and outputs are never shown.
2. An artificial neural network (ANN) is a sort of computer software, inspired by biological neurons. Biological brains are capable of solving difficult problems, but each neuron is only responsible for solving a very small part of the problem. Similarly, a ANN is made up of cells that work together to produce a desired result, although each individual cell is only responsible for solving a small part of the problem. This is one method for creating artificially intelligent programs.
3. Generative adversarial networks (GANs) are multiple artificial neural networks that work together to give better answers. One neural network is the *generative* network, and the other one is the *discriminative* network. The *generative* network will try to give an input to the *discriminative* network that will cause the useful network to give a bad answer. The *discriminative* network will then learn not to give a bad answer, and the *generative* network will try to trick the useful network again. As this continues, the *discriminative* network will get better and the percentage of wrong answers will fall.
4. Convolutional neural network(CNN) are a neural network that uses convolution instead of matrix multiplication. Convolutional networks were inspired by biological processes in that the connectivity pattern between neurons resembles the organization of the animal visual cortex. Individual cortical neurons respond to stimuli only in a restricted region of the visual field known as the receptive field. The receptive fields of different neurons partially overlap such that they cover the entire visual field. CNNs is a class of deep, feedforward ANNs, inspired by the Human Visual System, that have successfully been applied to analysing visual imagery and explicitly designed for complex feature extraction from two dimensional (2D) and three dimensional (3D) input volumes. Examples of CNN in computer vision are face recognition, image classification etc. It is similar to the basic neural network.
5. Recursive Cortical Network is the name of a software by the company Vicarious that is powered by Machine Learning. The system takes in sensory data, mathematics, and

biological plausibility in to make its assessment. For more details, feel free to read this article by Towards Data Science Inc.

6. Support Vector Machine :The objective of the support vector machine algorithm is to find a hyperplane (a separator) in an N-dimensional space(N — the number of features) that distinctly classifies the data points. To separate the two classes of data points, there are many possible hyperplanes that could be chosen. Our objective is to find a plane that has the maximum margin, i.e the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence.

The use of machine learning to aid CAPTCHA breaking can be split into the following categorization of approaches:

Aided by Machine Learning	Segmentation[1]	
	Non-segmentation	Reinforcement Learning
		Generative Adversarial Networks
		Recursive Cortical Networks
Driven by Machine Learning		Support Vector Machines
		Convolutional neural network

While segmentation had its day in dealing in CAPTCHA solving, the current analysis seems to be show that it is optimal to use Support Vector Machines in all phases of a CAPTCHA solving process (pre-processing, segmentation and post-processing) for best result for effort.

Improving CAPTCHAs

Several defensive design principles that can be done are for text-based CAPTCHAs:

1. Curved confusion lines: these lines causes machine learning algorithms to be confused as the the algorithms find it difficult to segment the image with the lines around.
2. Negative Kerning: By collapsing negative space between characters in a text-based captcha, each character will be close enough to seem to be in contact and make it difficult for the CAPTCHA solver to separate the characters for analysis.
3. Varying CAPTCHA length: By varying the length, the algorithm has to identify how many characters there are in an image. This has caused quite a number of solvers to fail.
4. Wave CAPTCHA: When arranged in a wave-like shape, the solvers again have issues identifying cut-off points to separate characters.
5. Randomize characters size and font type: this technique reduces the ease of learning by the solver
6. Rotate the characters: Deep learning techniques have been shown to fail after the characters/images are rotated.

In the end, a better way is to use other metrics (e.g. multiple sign-ups from single IP address) to evaluate the risk after the first Captcha (preferably not text-based) is done and bring up additional validation requirements if a threshold is met (e.g. SMS validation, different form of CAPTCHA).

References:

[1] Segmentation means splitting the CAPTCHA image into individual characters and dealing with them individually

SoK: Machine vs. machine – A systematic classification of automated machine learning-based CAPTCHA solvers, https://www.sciencedirect.com/science/article/pii/S0167404820302236

CWE-804: Guessable Captcha, https://cwe.mitre.org/data/definitions/804.html

Reinforcement learning. (2020, June 8). Wikipedia, The Free Encyclopedia. Retrieved 08:13, October 23, 2020 from https://simple.wikipedia.org/w/index.php?title=Reinforcement_learning&oldid=6980021.

Generative adversarial networks. (2018, March 10). Wikipedia, The Free Encyclopedia. Retrieved 08:14, October 23, 2020 from https://simple.wikipedia.org/w/index.php?title=Generative_adversarial_networks&oldid=6017392.

Artificial neural network. (2020, January 4). Wikipedia, The Free Encyclopedia. Retrieved 08:21, October 23, 2020 from https://simple.wikipedia.org/w/index.php?title=Artificial_neural_network&oldid=6764482.

Convolutional Neural Network , https://towardsdatascience.com/covolutional-neural-network-cb0883dd6529 .

Wikipedia contributors. (2020, October 21). Convolutional neural network. In Wikipedia, The Free Encyclopedia. Retrieved 08:41, October 23, 2020, from https://en.wikipedia.org/w/index.php?title=Convolutional_neural_network&oldid=984626156

Support Vector Machine — Introduction to Machine Learning Algorithms, https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47