DOI: 10.20472/IAC.2018.039.003

NIPUN BANSAL
Delhi Technological University, India

MUKUL SACHDEVA
DTU, India

TANISHA MITTAL
DTU, India

BREAKING AUDIO CAPTCHAS FOR IRCTC BOOKING AUTOMIZATION

Abstract:

CAPTCHAs are computer generated tests in the form of images, audios and object recognition that world can communicate easily and computer systems cannot. Internet sites present users with captchas to set apart human users from false computer programs, often referred to as bots. Their purpose is to obstruct attackers from performing automatic registration, online polling and other such actions. IRCTC, being the website to reserve tickets for Indian railways, one of the biggest railway network, has also employed both image and audio captchas for security purposes. However, the audio captchas used on the website are not effective in distinguishing between humans and bots. Most of the visual CAPTCHAs and some audio CAPTCHAs on different websites have been cracked using various methods of machine learning and we propound an identical idea to examine the security of audio CAPTCHAs on IRCTC website. In this paper, we show that our bot is able to break the IRCTC audio captchas with a success rate of 98%, 96.04% and 80.3% using three different models. Along with breaking the captcha, another python script written by us was able to automate the process of ticket booking. Thus, combining all of it into a single package could result in a system which would login and reserve tickets only by a single click. Travel brokers can easily use such a system for easy and fast booking of tatkal tickets which would lead to commercializing this activity for deriving huge profit from needy travelers.

Keywords:

Audio Captchas, Automatic Speech Recognition, IRCTC, Security, MFCC, Deep Learning

JEL Classification: L86, C80, D85

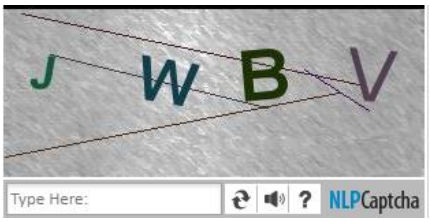
1. Introduction

Distinguishing computers from humans has ended up as a crucial topic for internet site security, as many services entrust on this security test to operate right. For instance, Gmail must inhibit misuse by automated spammers; Facebook must block the rapid rise of forged profiles used to transmit spam and IRCTC must block the abuse to prevent automatic booking of tatkal tickets by agents so that genuine people can reserve it.

Captchas (Completely Automated Public Turing tests to tell Computers and Humans Apart) are a type of challenge–response test used in computing to find whether the user is actually a human or some bot trying to perform an automated task. Image captchas usually take the shape of a twisted or rotated sequence of characters which are sometimes overlapping, whereas audio captchas are audio files with one or multiple speakers speaking characters at a different rate and intonation. They are simple for humans to understand, but comparatively troublesome for computers to process.

Need for Audio Captchas: It is easier and more natural for any user to look at a picture and type the contents i.e. use the image captchas.However, this idea of perpetual use of image captchas is based on a fairly large assumption: that all users have impeccable vision. The most common and approved quick fix to accessibility pitfalls of image captchas is to provide an added functionality of Audio Captcha along with image captchas, which allows users to request a pronunciation of the Captcha code. If, by any chance, the users are blind, color-blind, nearsighted, or have problem in reading the image captchas for any other reason – they can still figure out the captcha code by hearing the audio and then typing the code and access the Web without any hitch.

Figure 1. Image Captcha on IRCTC website



Source: https://www.irctc.co.in/eticketing/loginHome.jsf

Figure 2. Audio Captcha on IRCTC website

Source: https://www.irctc.co.in/eticketing/loginHome.jsf

Motivation Image captchas, though mostly used, enforce a limit on the millions of visually flawed people accessing the Web. Audio captchas were crafted to work out this problem. However, until presently, audio captchas have gotten less scrutiny in research as compared to image captchas. Recent research results exhibits that various accessible audio captchas are seriously ailing in security quality, as current machine learning methods are as of now genuinely progressed.In summation, most attacks on audio CAPTCHAs come at a comparatively low monetary value, as they involve merely a modest bit of training cases, rendering them somewhat enticing for attackers . Hence, we tested the security of audio captchas used on the IRCTC website.

Contribution IRCTC is the official website for booking tickets for Indian Railways. Our primary contribution through this system is the first ever attack, to our best knowledge, on IRCTC audio captchas. Our tool is able to crack the audio captchas with a success rate of 98%. This attack is significant in itself as Indian railways, being one of the largest railway network, caters to an approximate of 20 million people every day. To perform this attack, we have mainly performed two tasks -

1. **Scraping -** We wrote a python script to download the audio captchas from the IRCTC website.
2. **Breaking audio captchas -** Following the methodology suggested we constructed our own tool which is able to break the IRCTC audio captchas with an accuracy of 98%.

We also wrote a python script which can automatically book tickets on the IRCTC website for the users. Integrating this tool for breaking the audio captcha with automatic booking of tickets will make the entire process even simpler. The user will no longer have to worry about logging in or selecting trains. This further weakens the security of the website. This kind of a system can easily be used by a travel agent to book multiple tatkal tickets in few seconds, thus maximizing his profit and also depriving genuine and more needy travelers of the tatkal tickets.

Outline The remaining of the content is organized as follows: In Section 2 we walk through the steps involved in the process of breaking an audio captcha. Section 3 explains in detail how we constructed our classifiers. In Section 4, we display the

experimental results achieved by our system and analyze its performance. In the following Section 5, we present some suggestions that can be incorporated in the IRCTC audio captchas to reduce the attacks on the website. In Section 6, we provide some insights on the previous related works. In the end, we conclude in Section 7.

2. Cracking the Captcha Code

In this section, we explore how our tool can be used to break IRCTC audio captchas. IRCTC captchas are of varying length, 4 or 5, and comprise of digits from 0 to 9. Each digit is spoken by the same speaker, simply at different speed to develop the sound recording that the user hears. The pronunciation of some digits seemed to have been elongated to increase the difficulty of the captcha.

The entire process of decoding the audio captchas can be split into three phases, namely, segmentation, feature extraction and classification.

Segmentation This phase is used to break the audio into pieces which are to be recognized individually. Pre-processing of the audio is required to be carried out before segmentation of audio. In the preprocessing step, parts of audio irrelevant to the task, such as background noise, are removed. Since, captchas used on IRCTC do not have background noise, preprocessing was not necessary.

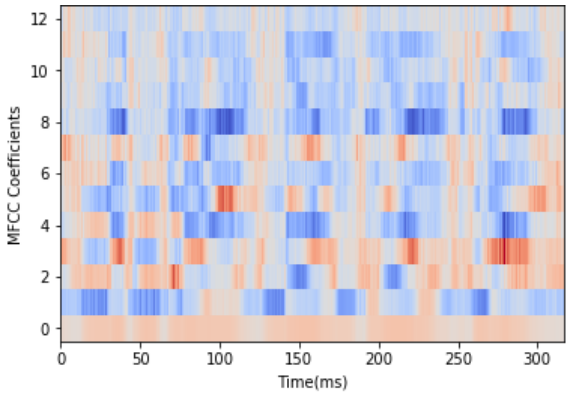
Segmentation is considered to be the most difficult stage of breaking audio captchas. Since the audio captchas we have worked on were varying in length, it was a more challenging task. To perform segmentation, each test captcha is divided into a number of segments of definite size, equal to the length of the captcha by detecting silence in between the digits and segmenting them using a python library. Since,the size of the feature vector formed from a segment ordinarily relies on the size of the segment and therefore using segments of fixed length allow each segment to be represented by a feature vector of the same size. To be competent in this phase requires a great deal of fine tuning of the parameters. We fine-tuned our splitting tool on the training set and selected the parameter using which the largest portion of our training set was being decomposed into its correct length. There was a little scope of improvement, but given the accuracy without it, it was not necessary.

Feature Extraction To crack the audio captchas, we extract features from the captchas and use various algorithms of machine learning to perform Automatic Speech Recognition on the segments of the audio captcha. There are many prevalent approaches for extracting features from audios, the most frequently used being MFCC.

MFCC, Mel Frequency Cepstral Coefficients, works quite akin to Fast Fourier Transform (FFT). MFCC converts an audio file to frequency bands just like FFT does but unlike FFT, MFCC makes use of mel- frequency bands, which are better for proximating the spectrum of frequencies that can be heard by humans[5]. Since all the digits in the captcha were spoken by a single speaker, we did not consider PLP

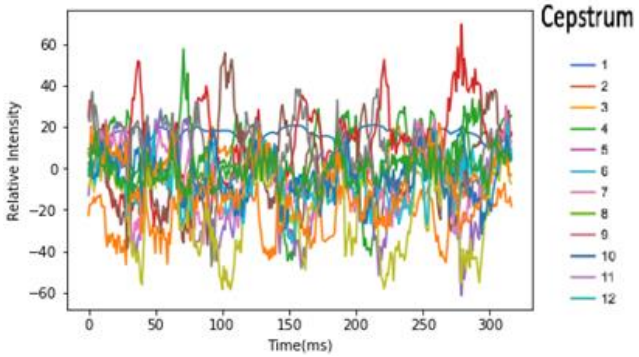
(perceptual linear prediction) feature representation which is used to extract speaker independent features from an audio. We used python library to calculate the MFCC features of the sound recordings.

Figure 3: Mel Frequency Cepstral Coefficients (MFCC) over time of a sample audio from training set. The 5 individual divisions are somewhat visible.



Source: Self Observed using a sample audio captcha

Figure 4: Mel Frequency Cepstral Coefficients (MFCC) over time of the same sample audio using different colors for different cepstrals.



Source: Self Observed using a sample audio captcha

Classification In the process of classification, individual segment is assigned a digit. It comes into the category of supervised classification and can be done by employing any suitable machine learning technique. At the end, the final prediction of the captcha is created by combining all classifications made for individual segments. We used a Convolutional Neural Network built in python to classify the audios.

3. Classifier Construction

The classification of audios uses the Deep Learning approach by implementation of a Convolutional Neural Network (CNN) for assigning a label to individual audios. The network has a series of convolutional , pooling and dropout layers. The final layer is a fully connected layer with the activation function **softmax** which classifies the audios into appropriate category and assigns it a number between 0 -9. We trained our data on three classifiers to draw a comparison between the performance of the different

activation functions.The classifiers differ only by the activation functions used in the hidden layers which are **sigmoid**, **tanh** and **relu**.

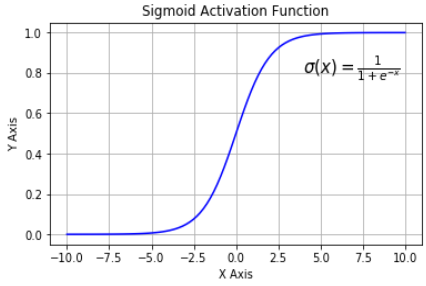
In the proposed method, we furnish a comparative study of the various activation functions that we used in the different layers of the three classifiers that we have built. Activation function can be simply defined as a function that decides whether the signal has to be passed from one layer to another in neural networks or not and introduces non-linear properties in the network. In our classifiers, we have principally applied the following activation functions.

i) Sigmoid : It takes a real-valued number 'x' as its input and converts it into a range of [0,1]. It transforms large negative numbers to 0 and large positive numbers to 1.Mathematically it is represented as

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad \text{- (Eq 1)}$$

Graphically, it is represented as:

Figure 5:Graphical representation of sigmoid (x)



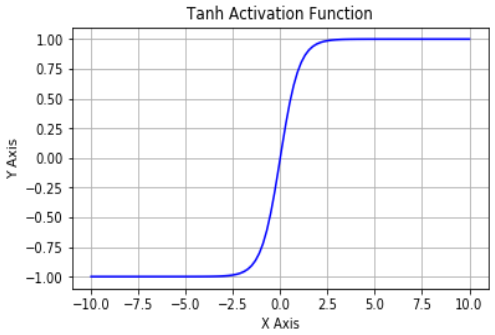
Source : https://www.learnopencv.com/understanding-activation-functions-in -deep-learning/

ii) Tanh : It is identical to Sigmoid and only differs in its output range which is [-1,1]. Unlike sigmoid, the outputs of tanh are centered around zero since the range is [-1,1]. Mathematically it is represented as

$$g_{\tanh}(z) = \frac{\sinh(z)}{\cosh(z)} = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}} \quad \text{- (Eq 2)}$$

Graphically, it is represented as:

Figure 6: Graphical representation of tanh (z)



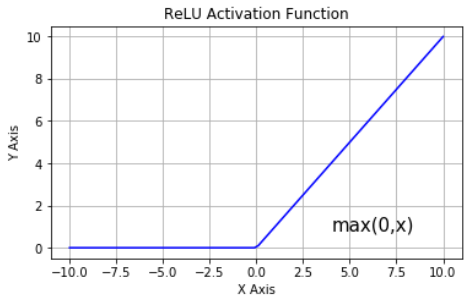
Source : https://www.learnopencv.com/understanding-activation-functions-in -deep-learning/

iii) Rectified Linear Unit (ReLu): It is chiefly implemented in hidden layers of neural networks. Mathematically, it is expressed as:

$$f(x) = \max(0, x) \quad \text{- (Eq 3)}$$

Graphically, it is represented as:

Figure 7: Graphical representation of ReLU (x)



Source : https://www.learnopencv.com/understanding-activation-functions-in -deep-learning/

iv) Softmax: It is mainly used for multi class classification.The i^{th} value of the N-dimensional output vector of the softmax function basically represent the probability of the input belonging to the i^{th} class. It maps:

$$S(\mathbf{a}) : \mathbb{R}^N \rightarrow \mathbb{R}^N$$

$$S(\mathbf{a}) : \begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ a_N \end{bmatrix} \rightarrow \begin{bmatrix} S_1 \\ S_2 \\ \cdots \\ S_N \end{bmatrix}$$

And the actual per-element formula is:

$$S_j = \frac{e^{a_j}}{\sum_{k=1}^{N} e^{a_k}} \qquad \forall j \in 1..N$$

-(Eq 4)

The loss function used after the fully-connected layer is Cross-entropy. Mathematically,

$$H(p,q) = -\sum_{x} p(x) \log q(x).$$

- (Eq 5)

Where p (x) is the wanted probability and q (x) is the achieved probability. The aim of this is to train our network in such a manner that the difference between the actual output and the achieved output i.e. error, is minimized.

4. Experimental Results and Performance Evaluation

The proposed method was tested with python on the dataset formed by scraping the audio captchas from the IRCTC website .

Table 1: Audios used in the training dataset

Number	Number of Audios
0	93
1	83
2	105
3	71
4	98
5	93
6	107
7	94
8	86
9	79

Source : Self Observation from training set

The most common and basic parameter to evaluate the performance of a classifier is **accuracy** which can easily be determined by Eq. (6)

$$Accuracy = \frac{TP}{TP + FP} \times 100\%$$

- (Eq 6)

where TP - true positives

FP - false positives

However, our tool may respond to some audio captcha and refuse to respond to others. For instance, since the length of IRCTC captchas is either 4 or 5, the system

would not respond to audios that have been segmented incorrectly in less or more than 4 or 5 individual audios. Thus, accuracy is not the only metric the system must be evaluated on. Therefore, following the approach of , we also determined the performance of the system on two additional parameters, namely, **coverage** and **precision.**

Coverage can be defined as the percentage of captchas that the system attempts to break. In other words, they are the fraction of captchas that have been segmented into the correct number of individual audios. Since the test data is same for all the three models, coverage is independent of the model used and depends on the segmentation technique used. **Precision,** on the other hand, is the fraction of captchas predicted correctly by the system from the captchas in the coverage set. According to, the captchas should be designed in such a way that ensures that “automatic scripts should not be more successful than 1 in 10,000” attempts,i.e. bots must have a precision of only 0.01%.

Table 2: Performance of our models illustrating the Acc (Accuracy), Cov (Coverage) and Pre (Precision)

The models used (differentiated on the basis of activation function used)		Parameters		
		Acc	Cov	Pre
	ReLU	98%	98%	100%
	tanh	96.04%	98%	98%
	Sigmoid	80.3%	98%	82%

Source :Self Observation from results

We have further depicted the results of the prediction made by the tool on the test set in the form of **confusion matrix**.

Table 3: Confusion Matrix of predictions made by model 1 (using ReLU as activation function)

		ACTUAL VALUES									
		0	1	2	3	4	5	6	7	8	9
PREDICTED VALUES	0	24	0	0	0	0	0	0	0	0	0
	1	0	25	0	0	0	0	0	0	0	0
	2	0	0	24	0	0	0	0	0	0	0
	3	0	0	0	22	0	0	0	0	0	0
	4	0	0	0	0	22	0	0	0	0	0
	5	0	0	0	0	0	23	0	0	0	0
	6	0	0	0	0	0	0	23	0	0	0
	7	0	0	0	0	0	0	0	17	0	0
	8	0	0	0	0	0	0	0	0	25	0
	9	0	0	0	0	0	0	0	0	0	24

Source: Self Observed using results

Table 4: Confusion Matrix of predictions made by model 2 (using tanh as activation function)

		ACTUAL VALUES									
		0	1	2	3	4	5	6	7	8	9
PREDICTED VALUES	0	24	0	0	0	0	0	0	0	0	0
	1	0	25	0	0	0	0	0	0	0	0
	2	0	0	24	0	0	0	0	0	0	0
	3	0	0	0	22	0	0	0	0	0	0
	4	0	0	0	0	22	0	0	0	0	0
	5	0	0	0	0	0	23	0	0	0	1
	6	0	0	0	0	0	0	23	0	0	0
	7	0	0	0	0	0	0	0	17	0	0
	8	0	0	0	0	0	0	0	0	25	0
	9	0	0	0	0	0	0	0	0	0	23

Source: Self Observed using results

Table 5: Confusion Matrix of predictions made by model 3 (using sigmoid as activation function)

		ACTUAL VALUES									
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
PREDICTED VALUES	**0**	24	0	0	1	0	0	0	0	0	0
	1	0	25	0	0	0	0	0	0	0	0
	2	0	0	22	0	0	0	0	0	1	0
	3	0	0	0	21	0	0	0	0	0	0
	4	0	0	0	0	22	0	0	0	0	0
	5	0	0	0	0	0	23	0	0	0	1
	6	0	0	0	0	0	0	22	0	4	0
	7	0	0	0	0	0	0	0	17	0	0
	8	0	0	2	0	0	0	1	0	20	0
	9	0	0	0	0	0	0	0	0	0	23

Source: Self Observed using results

It is notable, though, that previous related works had suggested that the present day captchas should be made more powerful by introducing captchas of varying length to mitigate the frequent attacks on audio captchas. However, our system was able to successfully break a captcha scheme which consists of audio captchas of length 4 and 5.

5. Mitigating the attack

We would like to suggest a few methods so as to mitigate the attack on the IRCTC website.

Limiting the number of downloads of captchas from an IP address: The most simple and straightforward technique for IRCTC to reduce the number of attacks on the website would be to slim down drastically the number of captchas that can be downloaded at a time by an IP address and the number of times a user can enter a wrong answer for the captcha code.

Though putting a limit on the number of downloads by IP address would restrict an attacker, he can easily attack with the same frequency as before using multiple IP addresses thus increasing only the attack cost. Therefore, this does not act as a viable solution for reducing attacks.

Adding background noise to the Captchas: Another effective strategy could be to add some noise in the background of the Captchas. A little noise in the background can make it difficult for the user to train the model and difficult for the machine to

recognize the captchas accurately. It will also make the process of segmentation troublesome.

Using other characters along with numbers:
Integrating numeric captchas with case-sensitive alphabets and some special characters like @, %, &, $ etc. can add to the strength of the captchas being used currently on IRCTC website.

Figure 8: Captcha with special characters and case-sensitive alphabet.

Source : https://fakecaptcha.com/

Increasing the pace of the audio: By increasing the speed at which speaker speaks the characters of the captcha, the task of segmenting the audio captcha can be made more onerous than the current method. Also, using varying length of silence or no silence between characters can make the task of providing a specific hard coded value for silence, to break the audio into parts, more difficult. It will further increase the degree of difficulty for an attacker to break the captcha.

Increased Vocabulary: Audio captchas can be presented as a more difficult test to be passed by pronouncing words instead of numbers and characters. Since the success of machine learning methods banks on the volume of training data we have, increasing the glossary and number of terms will result in less attacks as it will be difficult to collect enough training data.

6. Background and Additional Related Work

The most similar previous work focused on breaking the audio captchas of a specific website was Decaptcha: Breaking 75% of eBay Audio CAPTCHAs . The authors attempted to break the audio captchas on eBay and were successful in breaking 75% of the audio captchas used. They also compared their performance with a readily available speech recognition tool, Sphinx.

Visual captchas have been worked upon massively and there has been a record of rewarding attempts of cracking the captchas of famous websites.
For instance, in 2008, there was an attack on Microsoft's visual captcha system and it turned out to be 60% successful .

A considerable amount of work has also been done in exploring the audio captchas.

[5] proposed a methodology to break gmail's audio captchas. They examined the security of current audio captchas from popular internet sites like google.com, dig.com and an older variation of the audio captcha in recaptcha.net, by using machine learning algorithms like AdaBoost, SVM, and k-NN and obtained an accuracy of approximately 71%.

Hendrik Meutzner, Viet-Hung Nguyen, Thorsten Holz, and Dorothea Kolossar broke Google's reCAPTCHA with an accuracy of 52% and 63% using different models.

7. Conclusion

In this work, we showed how our tool can break 98% of IRCTC audio captchas.Attacks on the website can be slowed down by enforcing some limits on the number of captchas that can be scraped,by changing the captcha used on the website or by integrating scripts that detect the presence of bot and automated software on the website. An improved captcha scheme can be formed by incorporating some of the suggestions provided in the paper. Since the IRCTC captchas are varying in length, segmentation done is not hard coded and hence segmentation here is a major issue due to which the accuracy is not 100%. We plan to improvise our system by implementing better segmentation techniques.

References

Website from where data about usage and need of Captchas was collected. https://captcha.com/about/captcha-inc.html

Elie Bursztein and Steven Bethard. 2009. Decaptcha breaking 75% of eBay audio CAPTCHAs.In proceedings of the USENIX Workshop on Offensive Technologies (WOOT'09).

DANCHEV, D. Microsoft's captcha successfully broken.Blogpost http://blogs.zdnet.com/security/?p=1232, May 2008.

YAN, J., AND AHMAD, A. S. E. A low-cost attack on a Microsoft captcha. Ex confidential draft http://homepages.cs.ncl.ac.uk/jeff.yan/msn_draft.pdf, 2008.

Jennifer Tam, Jiri Simsa, Sean Hyde, and Luis von Ahn. 2008b. Breaking audio CAPTCHAs. In Proceedings of Advances in Neural Information Processing Systems (NIPS'15)

IRCTC website from where the dataset was collected https://www.irctc.co.in/eticketing/loginHome.jsf

Eli Bendersky's website providing information about the Softmax activation function and its properties. https://eli.thegreenplace.net/2016/the-softmax-function-and-its-derivative/

Understanding Activaton Functions in Deep Learning https://www.learnopencv.com/understanding-activation-functions-in -deep-learning/

K CHELLAPILLA, K LARSON, P. S., AND CZERWINSKI, M. Building segmentation based human-friendly, human interaction proofs. In 2nd Int'l Workshop on Human Interaction Proofs (2005), Springer-Verlag, Ed.

Hendrik Meutzner, Santosh Gupta, Viet-Hung Nguyen, Thorsten Holz, and Dorothea Kolossa. 2016. Toward improved audio CAPTCHAs based on auditory perception and language understanding. ACM Trans. Priv. Secur. 19, 4, Article 10 (November 2016)

Hendrik Meutzner, Viet-Hung Nguyen, Thorsten Holz, and Dorothea Kolossa. 2014. Using automatic speech recognition for attacking acoustic CAPTCHAs: The trade-off between usability and security. In Proceedings of the Annual Computer Security Applications Conference (ACSAC'14).