ISSN:0976-3031

Available Online at http://www.recentscientific.com

CODEN: IJRSFP (USA)

International Journal of Recent Scientific Research
Vol. 9, Issue, 4(F), pp. 25905-25911, April, 2018

International Journal of Recent Scientific Research

DOI: 10.24327/IJRSR

Research Article

CAPTCHA CRACKING TACTICS: IN GRAPHICAL TEXT BASED CAPTCHA

***Mir Aman Sheheryar and Zahoor Ahmed Najar**

Department of Information Technology,Central University of Kashmir,India

DOI: http://dx.doi.org/10.24327/ijrsr.2018.0904.1954

ARTICLE INFO

Article History:

Received 15th January, 2018
Received in revised form 25th January, 2018
Accepted 23rd March, 2018
Published online 28th April, 2018

Key Words:

Automated Bots, Turning Test, Cracking Artificial Intelligence.

ABSTRACT

A Captcha has its vital role for Web based Applications enjoyed over Internet. It serves as the security gateway and provides permission when it testifies parameters according to its requirements. CAPTCHA points out the difference between Robots (Automated Bot) and Human Beings. So that our system will not suffer illegitimate use. Continuous trends have evolved with the advancement of technology to defy the CAPTCHA Security system. With recent advancements in cracking techniques, some Approaches grew old and some attain attention with development. Herein we present detailed investigation of Cracking Approaches for Graphical Text Based CAPTCHA Techniques.

INTRODUCTION

After password the highly used authentication mechanism is the CAPTCHA mechanism where CAPTCHA is an acronym for "Completely Automated Public Turning Test To Tell Computer and Humans Apart"[1]. CAPTACH is based on "Turning Test" which works in accordance with HIP "Human Interactive Proof"[2]. Captcha provide secure compact Mechanism to protect and safeguard system against Automated Bots. After the evolution of CAPTCHA it provide compressive security mechanism for different types of services used over Internet. Though CAPTCHA provide coherent secure mechanism yet it faces hardships in terms to defend against cracking attempts. The attempt of cracking testifies the defense mechanism of CAPTCHA. CAPTCHA is the subset of "Artificial Intelligence" it runs in accordance to determine HCI, where HCI is "Human Computer Interaction" it explores the gateways for establishing contact between Machine and Human Being.

Day by day research is performed in the field of security and hence enormous amount of work is present on CAPTCHA. Time attacks were testified in [3], the authors examined the strength, frequency and explained detailed procedure to enhance the quality of strength and frequency. With relevant literature here in this paper we figured out trends to have healthy calculation for coming flow in cracking tactics for Graphical Text Based CAPTCHA.

Graphical Text Based CAPTCHA

The CAPTCHA was introduced in 2000 at "Carnegie Mellon University" in commercial scenario. In order to provide protection to 'Yahoo Chat room' Along with CAPTCAH introduction, cracking CAPTCHA by exploring took place. Through passing years many seminars, debates were held and often took place day by day in order to maintain stability in this platform.

The strength and protection of Graphical Text Based CAPTCHA merely lies in its Misrepresentation that involves Inference, Distortion. The main attribute of Graphical Text Based CAPTCHAare as under.

1. Broad intermixed character set, involving numbers and alphabets.
2. Special characters and symbols can also be the part of CAPTCHA along with numbers and alphabets.
3. Character string of variable length. This variation in length add stiffness against cracking.

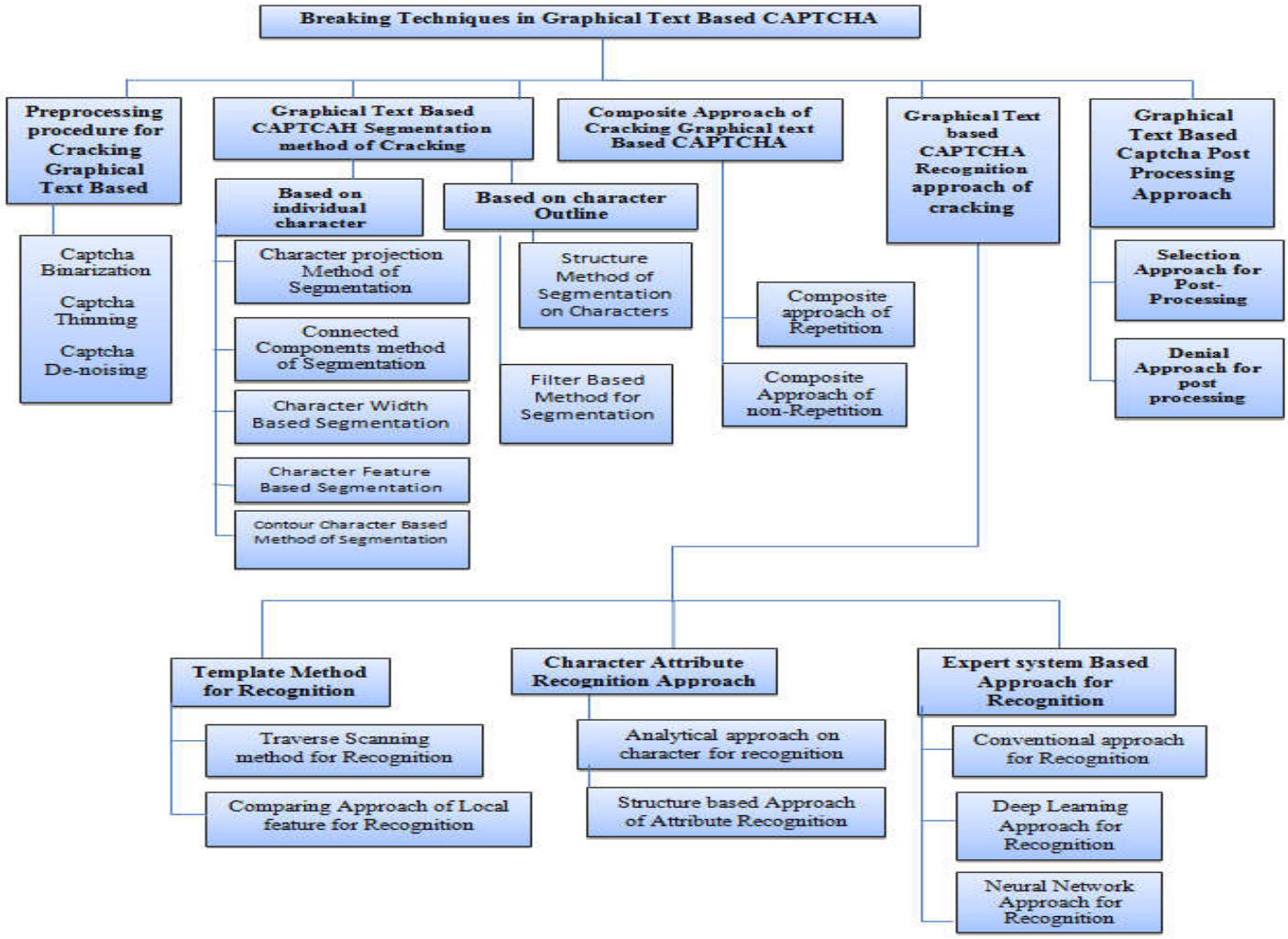
Corresponding author:* **Mir Aman Sheheryar
Assistant Professor Department of Information Technology, Central University of Kashmir, India

4. Color composition with respect to background should be same so as to withstand recognition percentage.

With the implementation of above mention attributes we can have CAPTCHA more protected against cracking possibilities.

Figure 1: showing Hierarchy of Breaking Techniques in Graphical Text Based CAPTCHA.

Figure 1 Hierarchy of Breaking Techniques in Graphical Text BasedCAPTCHA



Preprocessing procedure for cracking graphical Text Based Captcha

Preprocessing formulates the beginning state for defying Graphical Text Based CAPTCHA. In this procedure the CAPTCHA is firstly exploited for vulnerabilities to have soft bypass before subjecting Captcha to Segmentation and Recognition. The Preprocessing procedure involve CAPTCHA Binarization, CAPTCHA thinning and CAPTCHA Denoising apart from this some more procedures can also be employed depending upon the need for feasible solution.As in figure 2.

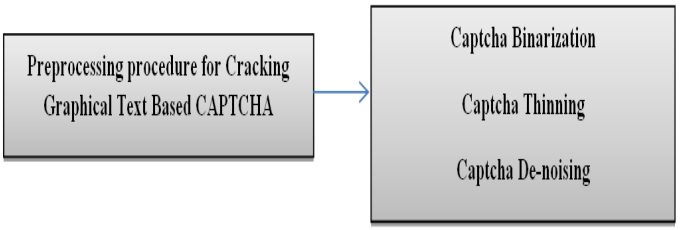


Figure 2 Preprocessing procedure for cracking graphical Text Based Captcha

CAPCTAH Binarization

In this the CAPTCHA is highlighted so as to eradicate background misrepresentation which is their, for making cracking tuff. In this CAPCTAH is determined mainly for global threshold, Dynamic threshold as used in sauvola[4], and otsu's method for threshold.

CAPTCHA Thinning

The Captcha boundary is highlighted for smooth processing. This also include subjecting CAPTCHA to Non iterative algorithm and iterative algorithm as Hilditch algorithm in[5], zhang and Suen algorithm in[6] explored CAPCTAH thinning ways. In CAPTCHA thinning the character space shall not vary before and after thinning.

CAPTCHA Denoising

It involves choosing appropriate technique for noise removal. As noise is present to misrepresent the CAPTCHA hence making it tuff to crack? Noise also results while CAPTCHA is subjected to Binarization. Different approaches for removal of noise are Domain spatial, Gibbs and Hough transform, Morphology based removal scheme, technique based on attached components, approach based on wavelet transformation.

Graphical Text Based CAPTCHA Cracking Advancements

Enormous amount of research is present for cracking CAPTCAH. Various researchers have designed different approaches to defy CAPTCAH. Mostly Segmentation and Non Segmentation Methods are adopted to defy CAPTCHA Security mechanism.

Segmentation Approach:this approach is adopted in phased manner depending upon task to solve. During this approach; character gap[7], connected region[8], vertical projection[9], connected regions[5] are performed on CAPTCHA for recognition scenario calculation is performed on the parameters to determine distortion evaluation[7], statistical character pixels[9], SVM[10].

Approach of cracking using segmentation on Graphical Text Based CAPTCHA are categorized under two scenarios (i) Based on Individual characters (ii) Based on character outline. As in Figure 3.

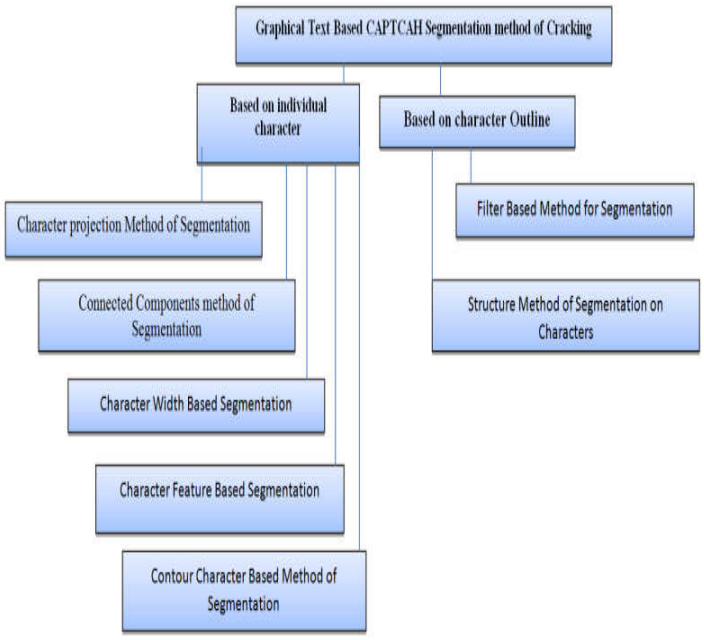


Figure 3 Graphical Text Based CAPTCAH Segmentation method of Cracking

Based on Individual characters

The Captcha is segmented and individual characters are obtained out of it. This approach involves character projection, connected components. Depending upon the outcome, we can use approaches on character width, character attributes and character contour.

Character Projection method of Segmentation

In this approach the precise feasible plane of character is determined by localizing among pixels subjected under different circumstances. This approach is applicable to determine tightly placed and loosely placed characters. Further connected components approach involveverticle, horizontal, projection guideline segmentation as in [1].

Connected components method of segmentation

This approach is employed for adjoint characters, in order to determine tilt keeping in view the skewness and tight coupling. In tight coupled character set image approach is not much worth. 90% of success rate was achieved in[11] where the researchers worked on MSN CAPTCHA using vertical projection and connected components. In their procedure CAPTCHA image string is partitioned and block of characters are obtained and highlighted. Block of characters is obtained bearing different colors. Finally vertical projection produces outcome that is feasible solution.

Character width based segmentation

This approach of segmentation is employed when CAPTCHA image are not segmented individually. Researchers in[12], used changing width to segment CAPTCHA. Each character segmented pertain recognition results of four that provide optimal solution. The widths of character were taken on average. Where in researchers in[10] avoid to use average width as standard instead they used dynamic programming to obtain optimal segmentation where minimum and maximum widths are taken into consideration.

Character feature based segmentation

The inside and outside attributes are used to determine CAPTCHA string feature. In side feature of characters were revealed in[13], where classification was performed that involve inside attributes.

Whereas researchers in [14], classified the outside feature with in the character set and provided" middle axis point separation" algorithm for CAPTCHA string. In this central pixel is used in background among two disjoint object pixels for points of segmentation.

Contour character based method of segmentation

In order to determine segmentation line contour character based method of segmentation is employed. Confidence interval was featured in research [15] to obtain segmentation line.

Based on character outline

Character outline segmentation involves structure based method of segmentation on character and filter based method for segmentation. The character outline segmentation focused on multiple characters than that of individual characters.

Structure based method of segmentation on character

Using this approach the segmentation is performed on component basis as the CAPTCHA image is discreted into black and white region, as discussed in[16] primarily black region is determined and white after that so as to figure out shared region to be identified easily.

Filter based method for segmentation

In this approach Captcha is disintegrated into four ways by image processing approach of convolution. Using "Gabor filters" as primarily introduced in[17].

Composite Approach of cracking Graphical Text based CAPTCHA

This approach involves combination of individual segmented characters as per their character set's so as to determine their recognition. Composite approach is followed by repetitive and non-repetitive approach of cracking. As shown in figure 4.

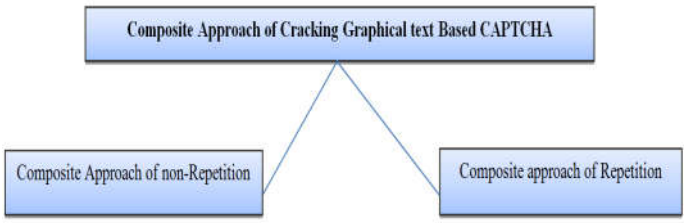


Figure 4 Composite Approach of cracking Graphical Text based Captcha

Composite approach of repetition

The character generated and character obtained through possible choice are more than actual character generated as in[17] and in order to determine each character is labeled from up, down, right and left.

Composite approach of non-repetition

The characters generated and obtained through possible choice equate each other. In this scenario character overlap stroke are used to make complete character in case of non- repetition as done in[18].

Graphical Text Based CAPTCHA Recognition Approach

The implementation of recognition approach include Template method for Recognition, Character Attribute Recognition approach and Expert system based approach for Recognition as in figure 5.

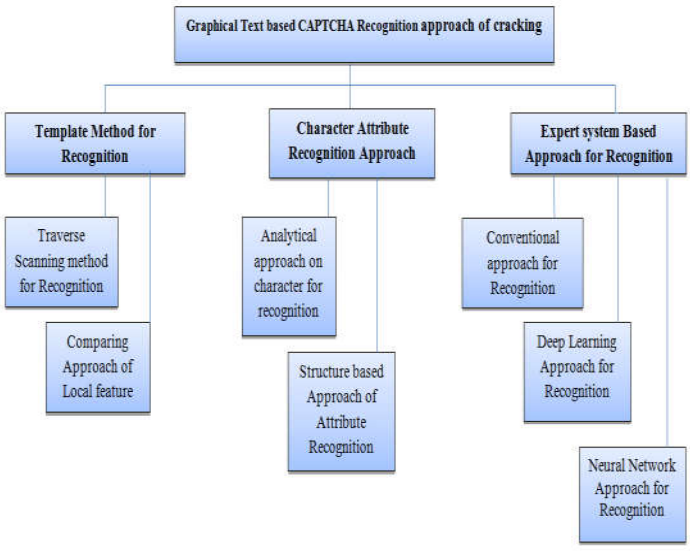


Figure 5 Graphical Text based CAPTCHA Recognition approach of cracking

Template method for Recognition

In this method of recognition similarity between thepixel characters are compared and most obtained i.e high similar is matched on the basis of local attributes and global property

Traverse Scanning method for Recognition

In this the similarity is obtained by region matching and correlation calculation. Firstly rough comparison is obtained and then exact comparison is made as figured out in[19].

Comparing Approach of Local feature for Recognition

In this the image is compared with the matching attributes of point set. Gimpy and EZ gimpy were cracked using local feature recognition by Malik in 2003[20] in order to achieve optimum outcome. This approach providehandy solution in "Shape Matching" and "Facial Recognition".

Character Attribute Recognition Approach

The layout of characters bedrock on the approach with which they are made, discrimination can be done by analyzing the behavior of their design. The character attribute recognition approach involves Structure Based Approach of Attribute Recognition and Analytical Approach on Characters for Recognition.

Structure based Approach of Attribute Recognition

Yahoo CAPTCHA was cracked using structure based approach of attribute recognition[21] involving detection of closed loop and direction of characters. This approach enables to analyze the physical demography of characters that include concavo-convex, loops, escalation point.

Analytical approach on character for recognition

In this approach the Boundary Attributes, Projection Attributes, Raw Snare Attribute are used to solve problems. This approach is mainly used for recognition as it stand worst for noise interference as discussed in[9] where CAPTCHA from Captchaservice.org were cracked with 100% success rate.

Expert System Based Approach for Recognition

These systems work on the behalf of the user by using the purpose based algorithms that provide robust outcome. The algorithms determine the CAPTCHA efficiently in order to crack the characters from it. Convolution approach, Deep Learning approach and Neural network approach for recognition are the three guided approaches for cracking CAPTCHA involving Expert System based approach for recognition. This approach is followed by training the systemfor number of iterations.

Conventional approach for Recognition

This Approach involvesthe use of classifiers like KNN and SVM. For obtaining closest neighbor sample KNN is used as in[17] KNN obtained highest success rate as compared to SVM, CNN. But was slower than CNN.

SVM discriminates the data so as to show separability using Hyperplane. In SVM Kernal function compares true attribute in nonlinear fashion into wide dimensional space as discussed by[10] where comparison of kernel functions were made using "RBF,POLY,LINEAR and SIGMOID".

Deep Learning Approach for Recognition

It involves the employment of CNN, LSTM-RNN,RNN. Deep Learning in recognition provides better result in recognition of Graphical text based CAPTCHA. Good recognition accuracy of CNN was described experimentally in[2,11,16,13] where recognition was obtained without attribute extraction.

Neural Network Approach for Recognition

The ability to work like human neuron form the bedrock foundation of Neural network approach for recognition. Due to the cognitive nature, Neural network is quite effective for CAPTCHA cracking. As shown in[22] where for cross entropy, back propagation neural network was used and as a result 53.3% of precession and 27,53% of precession was obtained on Captcha setsof MSN,eBay and Taboo.

Graphical Text Based CAPTCHA Post-Processing Approach

When the results obtained from previous approach needs further refinement. The obtained outcome is subjected to post processing where efficient outcome is obtained. The post processing approach comprises of selection approach and denial approach for post processing. As shown in figure 6.

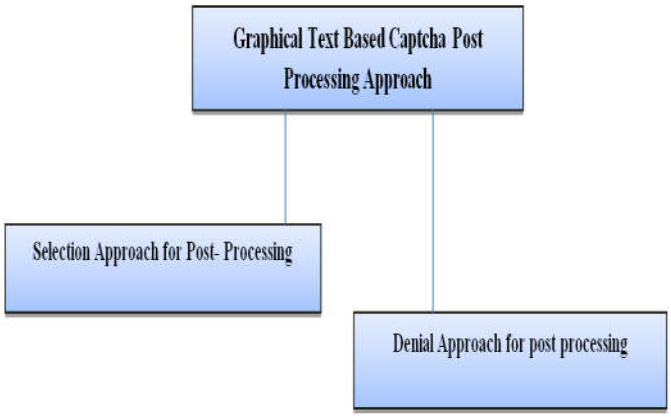


Figure 6 Graphical Text Based CAPTCHA Post-Processing Approach

Selection approach for Post-Processing

Using local and global optimization the selection approach for post processing is employed. In local optimization only confidence interval is taken into consideration for separate characters. The characters with high confidence interval are selected for high accuracy as in [23].

For collective character outcome global optimization is employed using traverse graph in[4] along with confidence interval. Researchers in[10,17] use dynamic programming in order to determine effective and accurate results.

Denial approach for Post-Processing

The approach of Denial is employed in order to verify tested sample behold the training set that is analyzed for recognition. The Denial approach ensures high CAPTCHA recognition.

Not much researchers have used the Denial approach for post processing. But string distance, Initial character, last character, confidence interval and multiple attributes were discussed in[22].

Comparative Analysis

On the basis of algorithm, approach employed and attribute selected.Various techniques have been evolved in order to determine the efficiency, advantages and disadvantage of Segmentation and recognition approaches. The comparison of segmentation approach and recognition approach are summed up in tabulated order as in table 1,2.

Table 1 Segmentation Approach Comparison

Segmentation Approach	Approach employed	Adherence	Tilt	Misrepresentation	Overlap	Description
Based on Individual character	*Projection* *Connected Components* *Width* *Attribute* *Outline*	Yes No Yes Yes yes	No Yes No No yes	No Yes No No Yes	No No No No No	Overlapped Limited overlapping Sever distort Notable
Based On individual Component	*Structure*	Yes	Yes	Yes	Yes	Easy to segment
	Filter	yes	Yes	yes	Yes	Complex

Strength and Weakness of Graphical Text Based CAPTCHA system were explored and discussed in[24] as summed up in Table3. Procedural approaches were employed so as to determine strength, weakness and drawbacks of Graphical Text Based CAPTCHA. In order to provide the path for designing and crafting sound Graphical Text Based CAPTCHA that will withstand Breaking attempts, so as to provide compressive security mechanism for web based applications used over internet.

Figure7 show the graphical result obtained after subjecting Chaac-CAPTCHA'H against cracking in[25]. It with stand cracking as compared to already existing CAPTCHA's, that include Gimpy, Ez-Gimpy, Secure image, Cryptography, Megaupload. Their work gave new CAPTCHA mechanism known as "Chaac-CAPTCHA" in two variants Easy and Hard variant and after experimental verification and calculation hard variant show 99% success rate to withstand cracking.

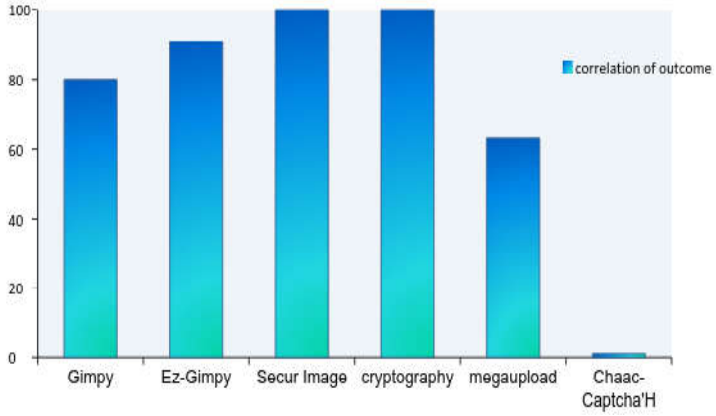


Figure 7 Result obtained in[25] showing Chaac-H withstand cracking Attack

Table 2 Recognition approach comparison

Recognition Approach	Attribute	Algorithm	Advantages	Disadvantages
Template Matching	*Global Property* *Local Attribute*	Travers search Shape Context	• Simple • Robust for Image scaling and Affine transformation	• Require large library • Rotation invariant.
Character Attribute	*Structure Attribute* *Analytical Attribute*	Character Structure Analytical Feature	• Detail sensitive. • Withstand noise interference.	• Application limited. • Distortion is high for noise interference
Expert system Based	*Template matching* *Conventional Approach* *Deep learning*	SVM,KNN BPNN,CNN CNN,RNN,LSTM-RNN	• Good Adaptability • Flexible structural design. • High recognition accuracy • Prevent gradient using time memory function.	• Limited to infinite samples. • Computational complexity. • Slow convergence rate. • Disappear time gradient • Inability of attribute extraction automatically.

Table3 Strength and Weakness of Graphical Text Based CAPTCHA System as discussed in [24]

CAPTCHA	Strength	Weakness	Drawbacks
Graphical Text based	1. Distorted text image is presented for the test to user. 2. Hollow, 3d, Opaque appearance of Text Image. 3. Tilt, skewness and deformation. 4. Similar background and foreground Colour.	1. High distortion makes it difficult for humans to read CAPTCHA Clearly. 2. Modern OCR's algorithms can achieve over 90% success in cracking.	1. In text images, user has someproblem to identify the correct textor characters. 2. Multiple fonts. 3. Font size. 4. Blurred Letters 5. Wave Motion. 6. Amount of Noise

CONCLUSION

With the detailed inquiry this research procures the advancements in cracking tactics of Graphical Text based CAPTCHA. Firstly we introduced different Graphical Text Based CAPTCHA Cracking Tactics fallowed by Hierarchy of cracking. Secondly we touted out detailed investigation for Preprocessing Approach, Segmentation Approach, Composite Approach, Recognition Approach and Post processing Graphical Text Based CAPTCHA approaches. After that we discussed the comparison in various research works and figured out the comparisons between segmentation approach and recognition approach depending upon attributes, algorithms, Advantages and Disadvantages. Finally we discussed the strength, weakness and drawbacks of Graphical Text Based CAPTCHA followed by experimental verification and calculation of hard variant Chaac-CAPTCHA as obtained in one of the research discussed which offer stiffness against cracking.

Future scope

Keeping in view the in-depth analysis for CAPTCHA cracking techniques we will be looking forward to devise the Recognition Engine. That will be Capable enough to Crack CAPTCHA depending upon the type of CAPTCHA and variants of CAPTCHA.

Acknowledgment

Highly thankful to Dr SaleemParveiz Mir for proof reading this paper and provide necessary motivation to procure this research.

References

1. L. Von Ahn, M. Blum, and J. Langford, "Telling humans andcomputers apart automatically", Communications of the ACM,vol. 47, no. 2, pp. 56.60, 2004.
2. K. Chellapilla and P. Y. Simard, ¡°Using Machine Learning toBreak Visual Human Interaction Proofs (HIPs), in Proceedingsof the Advances in Neural Information Processing Systems,pp. 265.272, ofAdvances in Neural Information ProcessingSystems, 2004.
3. N. Roshanbin and J. Miller, "A survey and analysis of currentCAPTCHA approaches", *Journal ofWeb Engineering*, vol. 12, no.1-2, pp. 001.040, 2013.
4. J. Sauvola and M. Pietiainen, "Adaptive document image binarization", *Pattern Recognition*, vol. 33, no. 2, pp. 225.236,2000.
5. C. J. Hilditch, "Linear Skeletons from Square Cupboards", *Machine Intelligence*, pp. 403.420, 1969.
6. T. Y. Zhang and C. Y. Suen, "A fast parallel algorithm forthinning digital patterns", *Communications of the ACM,* vol. 27,no. 3, pp. 236.239, 1984.
7. G.Moy,N. Jones, C.Harkless, and R. Potter, "Distortion estimation techniques in solving visual CAPTCHAs", in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004, pp. II23.II28, July 2004.
8. A. Bansal, D. Garg, and A. Gupta, ¡°Breaking a Visual CAPTCHA: ANovelApproach usingHMM,2008,https://pdfs.semanticscholar.org/3c2c/9af1e9a3b7095edaf8f205dfbadc30f 917fb.pdf.
9. J. Yan and A. S. El Ahmad, "Breaking visual CAPTCHAs with pattern recognition algorithms", in Proceedings of the 23rd Annual Computer Security

Applications Conference, ACSAC 2007, pp. 279.291, December 2007.
10. F. Jean-Baptiste and R. Paucher, ¡°The Captchacker Project,2009, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.800.3065&rep=rep1&type=pdf.
11. J. Yan and A. S. E. Ahmad, "A low-cost attack on a Microsoft Captcha", in Proceedings of the 15th ACM conference on Computer and Communications Security, CCS 08, pp. 543.554,USA, October 2008.
12. C. Hong, B. Lopez-Pineda, K. Rajendran, and A. Recasens, "Breaking Microsofts Captcha", 2015, https://courses.csail.mit.edu/6.857/2016/files/hong-lopezpineda-rajendran-recansens.pdf.
13. A. S. E. Ahmad, J. Yan, and M. Tayara, "The Robustness of GoogleCAPTCHAs", Computing Science Technical Report CSTR-1278, Newcastle University, 2011.
14. S.-Y. Huang, Y.-K.Lee, G. Bell, and Z.-H.Ou, "An efficient segmentation algorithm for CAPTCHAs with line cluttering and character warping", Multimedia Tools and Applications, vol.48, no. 2, pp. 267.289, 2010.
15. R. A. Nachar, E. Inaty, P. J. Bonnin, and Y. Alayli, "Breaking down Captcha using edge corners and fuzzy logic segmentation/ recognition technique", Security and Communication Networks, vol. 8, no. 18, pp. 3995.4012, 2015.
16. A. S. El Ahmad, J. Yan, and L. Marshall, "The robustness of a new Captcha", in Proceedings of the 3rd European Workshop on System Security, EUROSEC¡¯10, pp. 36.41,April 2010.
17. H. Gao, J. Yan, F. Cao *et al*., "A Simple Generic Attack on Text Captchas", in Proceedings of the Network and Distributed System Security Symposium, pp. 1.14, San Diego, Calif, USA, 2016.
18. B. B. Zhu, J. Yan, Q. Li *et al*., "Attacks and design of image recognition CAPTCHAs", in Proceedings of the 17th ACM Conference on Computer and Communications Security, CCS10,pp. 187.200, October 2010.
19. F. Dai, H. Gao, and D. Liu, "Breaking CAPTCHAs with second template matching and BP neural network algorithms", *International Journal of Information Processing and Management,* vol. 4, no. 3, pp. 126.133, 2013.
20. G. Mori and J. Malik, "Recognizing objects in adversarial clutter: breaking a visual Captcha", in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 134.144, June 2003.
21. H. Gao, W. Wang, Y. Fan, J. Qi, and X. Liu, "The robustness of connecting characters together CAPTCHAs", *Journal of Information Science and Engineering,* vol. 30, no. 2, pp. 347.369, 2014.
22. L. Zhang, L. Zhang, S.-G. Huang, and Z.-X. Shi, "A highlyreliable CAPTCHA recognition algorithm based on rejection",*ActaAutomaticaSinica,* vol. 37, no. 7, pp. 891.900, 2011.
23. C. Hong, B. Lopez-Pineda, K. Rajendran, and A. Recasens,"Breaking Microsoft's CAPTCHA," 2015, https://courses.csail.mit.edu/6.857/2016/files/hong-lopezpineda-rajendran-recansens.pdf.
24. Mir Aman Sheheryar, P. k Mishra and A. K Sahoo, "A Review on Captcha Generation and Evaluation Techniques", *ARPN Journal of Engineering and Applied Sciences,* VOL. 11, NO. 9, MAY 2016
25. Mir Aman Sheheryar, A. K Sahoo, "Chaac -Captcha: An Improvisation of Graphical based Captcha with Dynamic Random Misrepresentation for Discrimination BetweenHuman and Machine", *International Journal of Engineering Research & Technology* (IJERT), Vol. 6 Issue 05, May - 2017

How to cite this article:

Mir Aman Sheheryar and Zahoor Ahmed Najar.2018, Captcha Cracking Tactics: In Graphical Text Based Captcha. *Int J Recent Sci Res.* 9(4), pp. 25905-25911. DOI: http://dx.doi.org/10.24327/ijrsr.2018.0904.1954

****