

A survey of research on CAPTCHA designing and breaking techniques

Yang Zhang
School of Computer Science and
Technology
Xidian University
710071, Shaanxi, P.R.China
839805299@qq.com

Haichang Gao
School of Computer Science and
Technology
Xidian University
710071, Shaanxi, P.R.China
hchgao@xidian.edu.cn

Ge Pei
School of Computer Science and
Technology
Xidian University
710071, Shaanxi, P.R.China
424016505@qq.com

Sainan Luo
School of Computer Science and
Technology
Xidian University
710071, Shaanxi, P.R.China
490243478@qq.com

Guoqin Chang
School of Cyber Engineering
Xidian University
710071, Shaanxi, P.R.China
1033691094@qq.com

Nuo Cheng
School of Cyber Engineering
Xidian University
710071, Shaanxi, P.R.China
850428404@qq.com

Abstract—The Internet plays an increasingly important role in people's lives, but it also brings security problems. CAPTCHA, which stands for Completely Automated Public Turing Test to Tell Computers and Humans Apart, has been widely used as a security mechanism. This paper outlines the scientific and technological progress in both the design and attack of CAPTCHAs related to these three CAPTCHA categories. It first presents a comprehensive survey of recent developments for each CAPTCHA type in terms of usability, robustness and their weaknesses and strengths. Second, it summarizes the attack methods for each category. In addition, the differences between the three CAPTCHA categories and the attack methods will also be discussed. Lastly, this paper provides suggestions for future research and proposes some problems worthy of further study.

Keywords—CAPTCHA, text-based, image-based, audio/video-based

I. INTRODUCTION

CAPTCHA (Completely Automated Public Turing Test to Tell Computers and Humans Apart) [1] has been widely used as a security mechanism to prevent automated registration, spam or malicious bot programs. It usually generates and evaluates a test that is easy for humans to solve but difficult for computers [2]. If the success rate of solving a CAPTCHA for humans is higher than 90% and machines only achieve a success rate of less than 1%, this CAPTCHA can be considered a good one [6][7]. Therefore, it is widely accepted that a good CAPTCHA is not only usable but also robust.

Since CAPTCHAs were created, numerous variants have emerged. Existing CAPTCHAs can be classified into three categories: text-based, image-based and audio/video-based, as Fig. 1 illustrates. A text-based CAPTCHA is the earliest and most popularly used CAPTCHA scheme, especially based on English letters and Arabic numerals. Instead of typing characters through keyboards, image-based CAPTCHAs usually require users to understand the content shown in the CAPTCHA image and then conduct a mouse-based operation. Audio/video-based CAPTCHAs, as special CAPTCHA

schemes, are rarely used in current CAPTCHA systems but are still worth investigating.

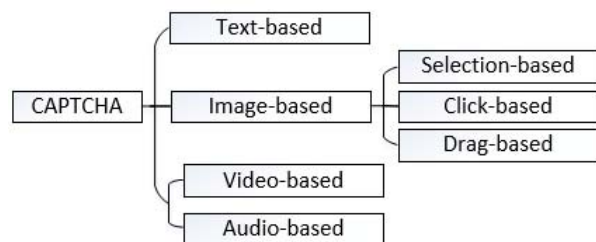


Fig. 1. Categories of existing CAPTCHAs

Researchers try to attack CAPTCHAs to verify their robustness and then identify which design features are good for security. Early attack methods were almost *ad hoc*, e.g., [5][62]. They have lost their efficiency with the evolution of CAPTCHAs. A generic attack followed a framework that mainly consisted of preprocessing, solution or other trivial modules that are becoming increasingly popular.

CAPTCHA technology is constantly developing, and related literature resources are gradually becoming abundant. Some previous literature [35][31][29][11][16][20][51][72] introduced CAPTCHAs at the time and investigated their security and usability, but they are either outdated or incomplete. It is necessary to summarize the emerging literature sources and conduct a more in-depth review on CAPTCHAs.

Our work reviews the relative research on CAPTCHAs. Not just the commonly used text-based and image-based, but also audio/video-based and some other newly emerging CAPTCHA schemes. For each CAPTCHA category, we discuss its usability and robustness and analyze its weaknesses and strengths. In addition, we summarize technical progress in attacking text-based and image-based CAPTCHAs. The detailed organization is shown in TABLE I. Finally, this paper

This work was supported by the National Natural Science Foundation of China under Grant 61472311.

compares three CAPTCHA categories as well as the attack methods and provides suggestions for other researchers. This work seems to be the most comprehensive literature review about existing CAPTCHA research, and some problems discussed in it are worth further study.

The remainder of this paper is organized as follows: Sections 2, 3 and 4 provide comprehensive introductions to text-based, image-based and audio/video-based CAPTCHAs, respectively. Section 5 summarizes the attack technology of CAPTCHAs in detail. Section 6 will compare different design mechanisms, attack methods and put forward helpful suggestions. Section 7 concludes the paper.

TABLE I. THE DETAILED ORGANIZATION

Design	Text-based	Anti-segmentation techniques	Hollow scheme
			CCT and overlapping
			Noise background
			Two-layer structure
		Anti-recognition techniques	Multi-fonts, Rotation and waving
			Large characters set
	Image-based	Selection-based	
		Click-based	
		Drag-based	
Attack	For text-based	Audio-based	Listener-model
			Speaker-model
		Video-based	
		Pattern matching-based	
		Pipeline-based	Segment
			Recognition
		End-to-end	
	For image-based	Preprocessing	
		Solution	

II. TEXT-BASED CAPTCHA

Text-based CAPTCHAs, the earliest and most deployed CAPTCHA type[7], typically asks users to recognize a distorted sequence that consists of English letters or Arabic numerals. Designers have applied various resistance mechanisms to improve their security, which can be classified

into anti-segmentation and anti-recognition mechanisms. TABLE II illustrates all of the resistance mechanisms.

TABLE II. EXAMPLES OF DIFFERENT SECURITY MECHANISMS

Security Mechanism		CAPTCHA examples
Anti-segmentation techniques	Hollow scheme	
	CCT and overlapping	
	Noise background	
	Two-layer structure	
Anti-recognition techniques	Multi-fonts, Rotation and waving	
	Large characters set	

Hollow: A main feature of hollow CAPTCHAs is to use contour lines to form connected characters with the aim of improving security and usability simultaneously, as the connected characters are hard to segment, but are easily seen by humans. Unfortunately, this mechanism is not as secure as people expected, research by Gao [17] used a generic method to combine segmentation with recognition to break a series of really hollow CAPTCHAs with success rates ranging from 36% to 89%.

CCT and overlapping: crowding characters together (CCT) and overlapping try to make segmentation more difficult by squeezing characters together. However, it may reduce user friendliness. For instance, reference [13] broke the Google CAPTCHA and reCAPTCHA [14] with success rates of 46.75% and 33%, respectively. A novel method was also presented in [15] to attack CCT-based CAPTCHAs, achieving success rates from 27.1% to 53.2%.

Noise background: The noise background mechanism hides the position of the characters. Regrettably, Google's reCAPTCHA, which uses Street View images, is broken by a method imitating the probability of a sequence proposed in [21]. In addition, Gao [19] used some image processing techniques iteratively to break PayPal's CAPTCHA.

Two-layer structure: A two-layer structure is a vertical combination of two single-layers CAPTCHA. The most critical issue in breaking this mechanism is segmentation, which cannot be solved by common segmentation methods. Gao [18] presented a novel two-dimensional segmentation approach to separate a CAPTCHA image along both vertical and horizontal directions and achieved a success rate of 44.6%.

Multifonts, Rotation, Waving: These three methods are designed to increase the diversity of each character, thereby increasing the number of features needed to identify each class by machine.

Large character set: A large character set makes the solution space much larger than that of traditional text CAPTCHAs. It usually chose Chinese, Japanese, etc.

Apart from the above work, reference [74] also proposed a new text-based user wuthentication technique which based a new algorithm.

III. IMAGE-BASED CAPTCHA

A. Selection-based CAPTCHA

The users are required to select the correct answers according to a hint for selection-based CAPTCHAs. It is the simplest form of an image-based CAPTCHA. Several examples of selection-based CAPTCHA are shown in Fig. 2.

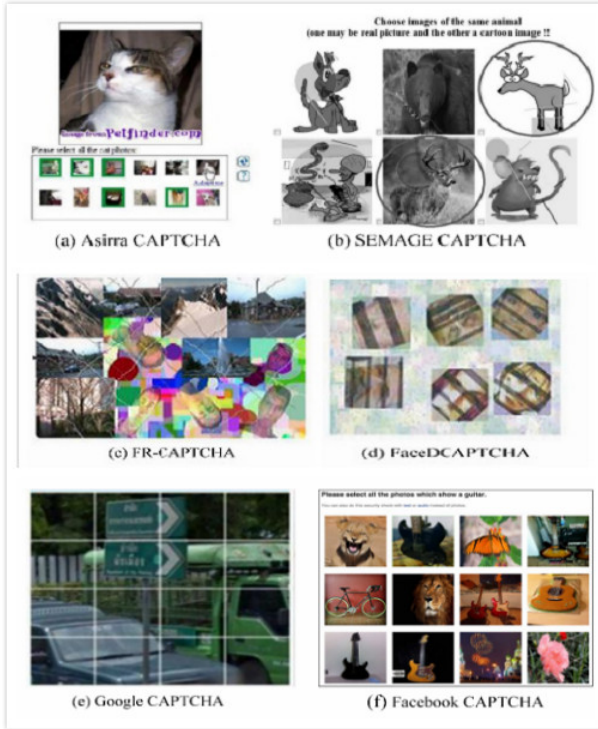


Fig. 2. Examples of selection-based CAPTCHAs

The **Asirra CAPTCHA** [27] asks users to select all the photographs with cats out of 12 photographs. This type of CAPTCHA was the first attempt at an image-based CAPTCHA. In contrast, the **SEMAGE** (SEmantically MAtching iMaGEs) CAPTCHA [32] asks users to select semantically related images from a given image set. This type of CAPTCHA exploits the ability of humans to accurately understand image content and establish semantic relationships between them. Avatar CAPTCHA is proposed in [59]. It asks users to identify avatar faces from a set of 12 grayscale images. The **FR-CAPTCHA** [33] and the **FaceDCAPTCHA**

[34] are face-based CAPTCHAs relying on human face recognition. **FR-CAPTCHA** asks users to select two face images of the same person. **FaceDCAPTCHA** requires users to distinguish the visually distorted real human faces among nonhuman face images. The **Google CAPTCHA** asks users to select all images with street signs or some specific object. The **Facebook CAPTCHA** asks users to select the corresponding images directly, according to a hint, from twelve images with different content. Recently, the work in [75] even construct a more sophisticated image CAPTCHA method by using semantic correlation to connect question keywords with answer choices.

Golle [28] proposed an SVM (Support Vector Machine) classifier to distinguish the images of cats and dogs in Asirra with an 82.7% success rate. In [36], Gao's team utilized OpenCV functions to detect faces in the FR-CAPTCHA, and four features were extracted from the faces to find the most probable pair. Reference [39] leveraged deep learning technologies to break an image reCAPTCHA and the Facebook CAPTCHA with success rates of 70.78% and 83.5%, respectively.

Selection-based CAPTCHAs are simple and convenient for users to operate. However, they have gradually become vulnerable due to the development of deep learning.

B. Click-based CAPTCHA

In 2008, Richard Chow et al.[23] first proposed the click-based CAPTCHA. It requires users to click characters in a complex background according to a short hint, as shown in Fig. 3. This CAPTCHA simplifies the user's operation, shortens the passing time and minimizes users' frustration.

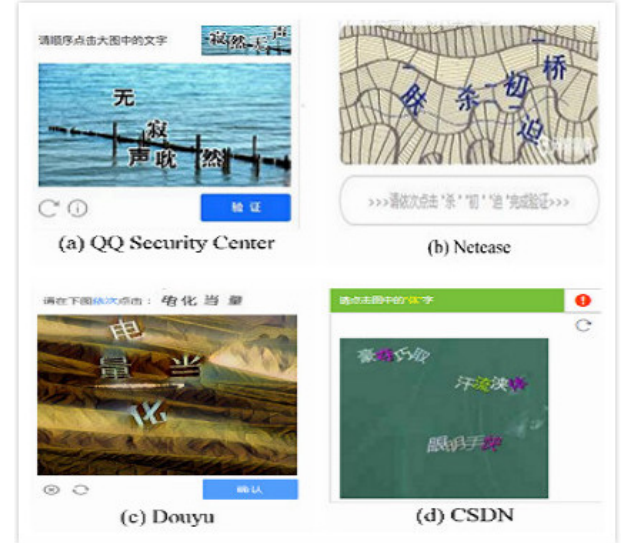


Fig. 3. Examples of click-based CAPTCHAs

In general, click-based CAPTCHAs have two defense mechanisms: anti-detection and anti-recognition. It is no longer a difficult task to recognize characters correctly with the development of machine learning. Therefore, almost all security mechanisms focus on preventing attackers from

correctly detecting characters. As shown in Fig. 3(c), the CAPTCHA uses the style transfer technique [24] to embed characters into the background to achieve the effect of hiding the characters.

Recently, a novel click-based CAPTCHA named VTT was proposed by Tencent (see Fig. 4). For a computer, it is difficult to understand the semantic information and analyze the image content as well as humans. In this regard, it seems a good design. However, state-of-the-art research on visual reasoning, such as [58] and [8], may break this scheme in the near future.

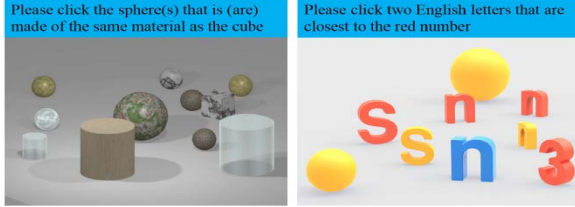


Fig. 4. Examples of VTT CAPTCHAs

C. Drag-based CAPTCHA

The drag-based CAPTCHA judges whether the user is a human through the mouse's track, speed and response time. The operation of a drag-based CAPTCHA is simple. Some drag-based CAPTCHAs are shown in Fig. 5.



Fig. 5. Examples of drag-based CAPTCHAs

The What's Up CAPTCHA, proposed by Google [37], is the first drag-based CAPTCHA. Users need to identify the upright orientations of randomly rotated images and adjust them to the correct position. Setting an image to an upright orientation is easy for humans, whereas it is difficult for bots. It is worth noting that images used in CAPTCHA must be manually filtered from samples that do not contain clear directional information. After this development, GeeTest proposed the first version of a slider CAPTCHA. Users must drag a slider along a line to the specified position continuously. Inspired by this, the VAPTCHA appeared. It

asks users to draw a trajectory with a mouse according to an arrow trajectory embedded in the background.

In fact, early drag-based CAPTCHAs judged the legitimacy of users only by measuring the speed of their operation. Therefore, it can be easily imitated. The later drag-based CAPTCHAs usually incorporate some user background data analysis. At present, this seems to be the future development trend for CAPTCHAs.

IV. AUDIO/VIDEO-BASED CAPTCHA

A. Audio-based CAPTCHA

This CAPTCHA is usually considered an alternative to a visual CAPTCHA in the case of visually impaired users [44]. Users in most audio-based CAPTCHAs play the role of listeners, and they are required to complete the specified challenge based on what they have heard. A spoken CAPTCHA system was introduced in [45]. This system converts a selected word into speech using a Text-To-Speech (TTS) system, then plays the sound clip to users and asks them to say the word. In 2012, the SoundsRight audio CAPTCHA (Fig. 6(a)) provided in [48] asks users to identify a specific sound, such as the sound of a bell or a piano. This work has increased the success rates in audio Captchas from less than 50% to over 90% for blind users. Meutzner et al. presented a new type of audio CAPTCHA [50] that uses additional nonsense speech sounds that are confusing for speech recognizers, while being less critical for human listeners. In 2016, they also proposed a nonspeech audio CAPTCHA [61], which is entirely based on the classification of sound events mixed into an environmental scene. Moreover, the HuMan CAPTCHA designed in [60] asks users to answer the presented questions by combining ambient noise and common sense knowledge. There is another type of audio-based CAPTCHA in which users play the role of speakers and are required to pronounce rather than simply listen. For instance, Gao et al. [47] proposed a new sound-based CAPTCHA (Fig. 6(b)) that exploits the differences between a human voice and a synthetic voice. A user is required to read out a given sentence rather than listening an audio file.



Fig. 6. Examples of audio-based CAPTCHAs

Attack and defense always go together. A two-stage attack for the listener model can always obtain a good attack result. In detail, the audio-based CAPTCHA is segmented into several regions regarding the location of each spoken word first. Then, the regions are recognized by automatic speech recognition programs. Tam et al. achieved success rates of up to 71% (Google Audio CAPTCHA, reCAPTCHA Audio CAPTCHA, Digg CAPTCHA)[44]. Bursztein et al. achieved success rates of 45%, 49% and 83% on the CAPTCHAs of Yahoo, Microsoft and eBay, respectively[49]. Some

researchers have even proposed that most of the digit-based audio CAPTCHAs are successfully broken with success rates between 50%-90% [44][52][54].

B. Video-based CAPTCHA

In video-based CAPTCHAs, users are provided a video file, and they should choose one option that best matches the video. Motion CAPTCHAs (Fig. 7(a)) proposed in [46] asks users to select the sentence that describes the motion of the person in the video. Rao et al. [55] proposed a video CAPTCHA (Fig. 7(b)) based on advertisement recognition. Both CAPTCHAs need users to select from the options provided. That is, these schemes can be broken by random guessing. Contrastly, reference [42] presented a video-based CAPTCHA (Fig. 7(c)), which asks users to type three words that best describe a video.

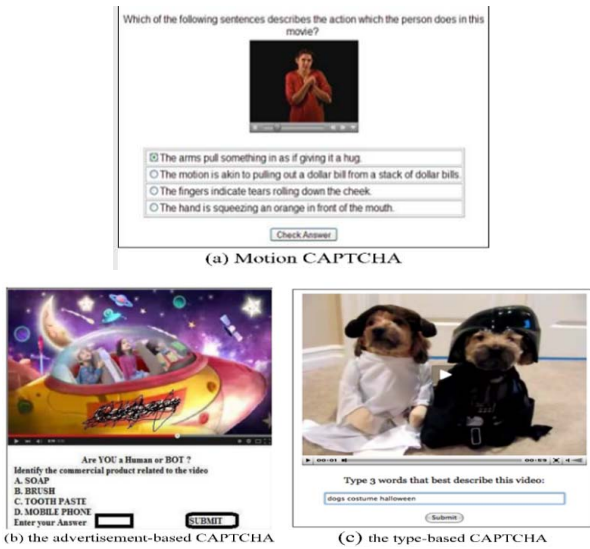


Fig. 7. Examples of video-based CAPTCHAs

Due to recent advances, bandwidth and video analysis technology no longer limit the development of video-based CAPTCHAs. In 2015, Sano et al. first used HMM-based (hidden Markov model-based) attacks to successfully attack a video-based reCAPTCHA from Google with a 31.75% success rate [65].

V. RESEARCH ON ATTACK METHODS

In general, the attack methods for different CAPTCHAs have different strategies. Fig. 8 shows the whole framework.

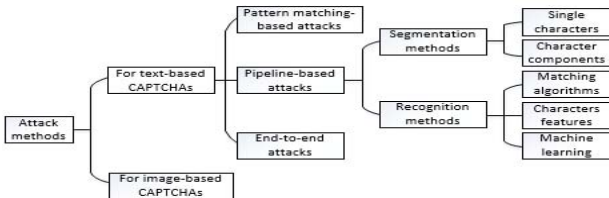


Fig. 8. The framework of breaking technology

A. Attack methods for text-based CAPTCHAs

1) Pattern matching-based attacks

In the first stage, a traditional pattern matching-based attack is used. This earliest methods to break CAPTCHAs are more straightforward. Back in 2003, Mori et al. [5] used a method based on shape context matching to break EZGimpy, which was used by Yahoo and Gimpy CAPTCHA with success rates of 92% and 33%, respectively. EZGimpy and Gimpy were also broken by a correlation algorithm and a direct distortion estimation algorithm [62].

In fact, the early attacks on CAPTCHAs are mainly based on artificial feature extraction, which is different from the later classifier. The feature extraction method is still based on pixel-level matching. These methods are almost always *ad hoc*, which are not only costly but also have limited effects. This method has no fixed form but the idea of pattern matching. The lack of good breaking technology leads to the slow development of CAPTCHA mechanisms. However, with the increase of the single-character recognition rate, the technology is developed gradually at this stage.

2) Pipeline-based attacks

In the second stage, the attack methods always follow a pipeline with segmentation and recognition. Before that, some alternative preprocessing methods are chosen before the segmentation stage. The main purpose of these preprocessing methods is to eliminate noise and highlight the information part. Grayscale and binarization are the most common two methods. In addition, erosion and dilation algorithms are two effective methods to remove noise. The opening operation (dilation after erosion) is useful for removing small noise elements, and the closing operation (erosion after dilation) is used to fill the tiny breaches. Some filtering methods can also be used, such as the mean filter, which can address salt-and-pepper noise [3].

a) Segmentation

- Segmentation methods based on single characters

Uniform cutting: This method is appropriate for CAPTCHAs in which the width of characters is similar and their distribution is uniform. This is a simple but limited approach.

Feature extraction: This method can be used for characters such as alphabetic characters and Arabic numerals. It uses some character features, such as circles and dots. These methods will design some algorithms to detect characters that contain a dot shape such as in “i” and “j”, a loop shape such as in “a”, “b”, a cross shape such as in “t” and “f” and so on.

Projection: Computing the character projection histogram is an effective method for segmenting individual characters. It is suitable for nonoverlapping characters. The direction of the projection can be vertical and horizontal, as well as the combination of these two directions. Previous work in [10] used the vertical projection method for segmentation. (see Fig. 9)

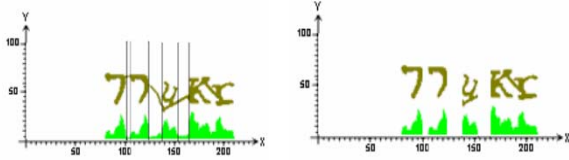


Fig. 9. An example of vertical projection segmentation

CFS: CFS (color filling segmentation) is used to detect each connected component using the CFS algorithm [10]. The CFS algorithm contributes to further segmentation by detecting objects that cannot be segmented by projection. Previous work in [17] uses color filling to detect the connected components shown in Fig. 10 and [10] uses the color filling method in the background for loop detection.



Fig. 10. The CFS method in [10]

- Segmentation methods based on character components

Filter: This method is appropriate for a wide range of text-based CAPTCHAs. It extracts character components in a CAPTCHA image with filters of different orientations and then combines these adjacent components in different ways to form individual characters. In [69], Gao et al. use Log-Gabor filters with four different orientations. (Fig. 11)

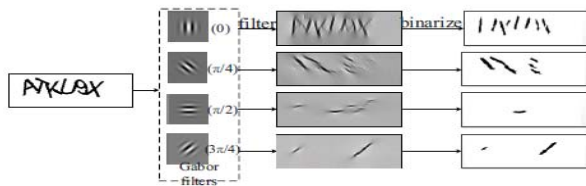


Fig. 11. Extracting character components with a Gabor filter.

Character structure: In [17], the contour lines of characters constitute some close parts, as shown in Fig. 12. According to this close structure, after color filling and noise component removal, the strokes of the character are segmented.



Fig. 12. Segmentation in [17]

b) Recognition

- Recognition methods based on matching algorithms

Matching algorithms compare the similarity of each pixel between characters. It consists of two algorithms: template matching and shape context. Methods based on template matching compute the confidence level between the images

that need to be recognized and the template images[70]. The shape context is based on the description of the object's contour sample points[5].

- Recognition methods based on characters features

This scheme relies on differences in characters, such as the shape. For the shape feature, Gao et al.[41] analyzed characters in different kinds of CAPTCHAs and proposed a method consisting of the approaches 'cut head and tail (CHT)', 'the guide lines principle' and 'the loop principle' to break CAPTCHAs from Yahoo!, Baidu and Google. For the statistical features, Jeff Yan[43] broke the CAPTCHA presented in [56] with a success rate of higher than 90% by counting the number of pixels per character and using a dictionary attack. By using the distinct pixel count for each letter, they simultaneously broke most visual CAPTCHAs with a near 100% success rate in [68].

- Recognition methods based on machine learning

Both of the methods mentioned above are based on the characters' features, which are vulnerable to noise disturbance. Therefore, as CAPTCHAs become more complex, these two recognition methods will not be applicable. In recent decades, more and more visual problems can be solved by machine learning. Therefore, there is no doubt that a CAPTCHA, a type of visual problem, can also be solved by this way.

In 2016, Gao [69] proposed using a kNN (k-nearest neighbor) as a recognition engine. The Captchacker Project presented in [66] exploits the potential of SVMs to break visual CAPTCHAs. The work in [28] describes a classifier that combines two SVMs to break the Asirra CAPTCHA. Currently, CNNs (convolution neural networks) have become a powerful tools for classifying characters with a high success rate. Regardless of the type of security mechanisms used, a CNN always has a high tolerance. Gao [19] et al. segmented a sequence into several characters and then used a CNN for recognition to break a series of CAPTCHAs. In addition, Kopp et al. Reference [30] were able to break eleven CAPTCHA schemes using a CNN for localization as well as recognition with a success rate higher than 50%. In addition, deep learning is used in [57][39] as well. The work in [73] even presented a novel method of selective learning confusion class for text-based CAPTCHA recognition.

3) End-to-end attacks

Due to the complexity of multistage processing and the powerful classification capability of deep learning, the methods of end-to-end attack have received renewed attention from researchers. It is not only simple and efficient but it also greatly shortens the prediction time.

Liang et al.[63] were the first to present an algorithm using an RNN (Recurrent Neural Network) to recognize CAPTCHAs. To improve the reliability of the recognition results, a new algorithm based on an SVM was proposed. Subsequently, a two-dimensional LSTM-RNN was used to recognize text-based CAPTCHAs in [64]. To further improve the result of the breaking algorithm, they also proposed a novel algorithm based on the multi population genetic algorithm. In addition, the work in [21] used a method

imitating the probability of a sequence by factoring it and trained a CNN to evaluate these probability factors to break a reCAPTCHA. Recently, George [22] presented a generative compositional model that learns from a small dataset. It fundamentally breaks the defense of many text-based CAPTCHAs and is even better than some CNN models. The technology of GAN is also used in [72] which considers the benefits of merging the role of designer/attacker into the one system and speeds up the whole adversarial issue inherent in CAPTCHA.

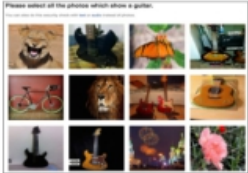

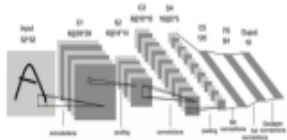









B. Attack methods for image-based CAPTCHAs

For image-based CAPTCHAs, the attack method usually adopts the two-stage method, which includes preprocessing and solution. In the preprocessing stage, there is the segmentation of small pictures, the location of the slider, the location of the gap and so on. To locate the object, some traditional image processing techniques, such as binarization and projection, are used. In addition, some object detection methods using deep learning can accurately locate the characters. After that, in the solution stage, the classification of the small pictures, the movement of the slider, and the

filling of the gap are carried out to obtain the final results. The use of an SVM, a kNN, a CNN, etc. are all optional. Some examples are given in Table III.

In TABLE III, the first line is the selection-based CAPTCHA, which is the simplest form of an image-based CAPTCHA. Due to the regular form and fixed arrangement of sub images, it is generally simple to extract each small picture and then corresponding labels should be provided for each sub image. The work in [39] also proposed a novel low-cost attack that leverages deep learning skills for the semantic annotation of images. Similarly, the work in [57] even constructed an association graph to break the image CAPTCHA of China's railroad system. They achieved a 230-dimensional latent space for the phrase first and then computed the likelihood of each of the eight images and the phrase by an association graph. The system broke 77% of the image CAPTCHAs in two seconds on average. Recently, some image classification services based on deep learning have gradually become open sources. However, using online image classification API's to directly attack selection-based CAPTCHAs is a more efficient method.

TABLE III. DETAILS OF ATTACK METHODS FOR IMAGE-BASED CAPTCHA

	CAPTCHA	Pre-processing	Solution
Select-based		 segmentation	 classification
Click-based		 location	 solution
Drag-based		 location	 solution
		 location	 solution

Increasingly high attack rates have led to the abandonment of selection-based. Instead, click-based CAPTCHAs have come more into view. However, the good times will not last long. The proposed object detection algorithms greatly weakened its robustness. For example, the RCNN [9][40][26],

R-FCN [53], YOLO [67][25][12] and SSD [4]. The work in [38] used the online OCR API, which is based on these object detection algorithms, to successfully break the click-based CAPTCHAs from Geetest, Tencent and Netease.

With the development of deep learning technology, a type of drag-based CAPTCHA is proposed, which usually requires users to drag a slider or mouse to complete the track verification of one or several segments. The difficulty of this form lies not only in the matching of trajectories but also in the analysis of the background behavior data of users during the verification process and the judgment of user legitimacy by some machine learning algorithms. Of course, there is no such thing as absolute security. Reference [38] proved that, as the first to detect the puzzle region by comparing the background and its source image, and then mimicked human behaviors by leveraging four simulation functions. They successfully attacked the drag-based CAPTCHA from Geetest, Tencent and Netease with more than a 96% success rate.

VI. DISCUSSION

A. Comparison between three kinds of CAPTCHAs

The text-based CAPTCHA is the earliest and most deployed CAPTCHA. The security of them can be increased by adding different font faces, noise and so on. However, compared with other CAPTCHAs, these schemes require more complex operation, and many of them are no longer safe.

The image-based CAPTCHA is the most diverse type. Easy operation is the advantage of image-based CAPTCHAs. Although some image-based CAPTCHAs are vulnerable to deep learning attacks, it still has a large development space.

Audio/video-based CAPTCHAs are not common in the real world. These two types of CAPTCHAs require higher bandwidth and take more time for users. More importantly, most of them can be easily broken. Based on our general experience, we summarize our analysis in TABLE IV.

TABLE IV. COMPARISON BETWEEN THREE TYPES OF CAPTCHAs

	Safety	Usability	Mobile deployment
Text-based CAPTCHA	Low	Middle	Middle
Image-based CAPTCHA	Middle	High	High
Audio/video-based CAPTCHA	Middle	Middle	Middle

B. Comparison between the three stages of attack methods for text-based CAPTCHAs

Pattern matching-based attacks can be used for the early simple CAPTCHA. However, the disadvantage is that for each different type of CAPTCHA, manual guidance is needed to extract the features, which is difficult to perform.

Pipeline-based attacks are commonly used. However, the multistage processes are time consuming, which seriously affects the breaking speed. This is also the main reason why this method is not suitable for real-time attacks.

End-to-end attacks usually use deep learning techniques to identify the entire CAPTCHA. The disadvantage is that the

network is difficult to train and the hardware requirements are high. Based on our general experience, we summarize our analysis in TABLE V.

TABLE V. COMPARISON OF THE THREE STAGES OF ATTACK METHODS

	Need training process	Accuracy rate	Recognition speed
Pattern matching-based attacks	No	Low	Slow
Pipeline-based attacks	Yes	Middle	Faster
End-to-end attacks	Yes	High	Fastest

C. Suggestions

1) Suggestions for designers

Deep learning technology: Deep learning technology is not only a way of breaking but also a tool to enhance the security of CAPTCHAs. For example, adversarial examples and neural style transfer may point out a new directions for the design of CAPTCHAs.

Some hidden information: In addition to focusing on whether the challenge is solved successfully, some hidden information such as pass time, pass speed and the operation track is useful to distinguish human and computers as well.

The ability of understanding: The ability of semantic understanding deserves more attention. We should make full use of people's excellent semantic understanding to enhance the robustness of CAPTCHAs. Besides, some other advanced human abilities can also be used, such as associative power.

2) Suggestions for attackers

Better network for end-to-end attacks: End-to-end attack is a new direction for CAPTCHA breaking. However, the combination brings difficulties in the training step. We expect a better and lightweight network

Fewer training datasets to attack: Although deep learning can extract effective features, the requirement for a large number of annotated datasets is indeed a problem that cannot be ignored. We look forward to a model that is less dependent on the training dataset.

More generic networks for different CAPTCHA mechanisms: For each type of CAPTCHA, we need to train a separate network to break it. The cost of labeling datasets is very high, it is not cost-effective. We need a more generic network, which can be migrated to different mechanisms.

VII. CONCLUSIONS

Based on in-depth analysis, this paper reviews the security mechanisms of three kinds of CAPTCHA, as well as the attack methods. We also discuss the difference between the three types of schemes and attack methods for text-based. Finally, meaningful suggestions are proposed for designers and attackers. We hope our work will highlight new directions

of CAPTCHA study and provide ideas and inspiration to other researchers.

REFERENCES

- [1] Von Ahn, L., Blum, M., & Langford, J. (2004). Telling humans and computers apart automatically. *Communications of the ACM*, 47(2), 56-60.1
- [2] Von Ahn, L., Blum, M., Hopper, N. J., & Langford, J. (2003, May). CAPTCHA: Using hard AI problems for security. In *International Conference on the Theory and Applications of Cryptographic Techniques* (pp. 294-311). Springer, Berlin, Heidelberg.
- [3] Gonzalez R C, Woods R E, Eddins S L. *Digital Image Processing Using MATLAB: AND "Mathworks, MATLAB Slim SV 07"*[M]. Prentice Hall Press, 2007.
- [4] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[J]. 2015:21-37.
- [5] Mori, G., & Malik, J. (2003, June). Recognizing objects in adversarial clutter: Breaking a visual CAPTCHA. In *Computer Vision and Pattern Recognition*, 2003. Proceedings. 2003 IEEE Computer Society Conference on (Vol. 1, pp. I-1). IEEE.
- [6] Chellapilla, K., Larson, K., Simard, P., & Czerwinski, M. (2005, April). Designing human friendly human interaction proofs (HIPs). In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 711-720). ACM.
- [7] Yan, J., & El Ahmad, A. S. (2008, July). Usability of CAPTCHAs or usability issues in CAPTCHA design. In *Proceedings of the 4th symposium on Usable privacy and security* (pp. 44-52). ACM.
- [8] Krishna, R., Chami, I., Bernstein, M., & Fei-Fei, L. (2018). Referring Relationships. *arXiv preprint arXiv:1803.10362*.
- [9] R. Girshick, J. Donahue, T. Darrell, J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.
- [10] Yan, J., & El Ahmad, A. S. (2008, October). A Low-cost Attack on a Microsoft CAPTCHA. In *Proceedings of the 15th ACM conference on Computer and communications security* (pp. 543-554). ACM.
- [11] Bhalani, S. D., & Mishra, S. (2015). A survey on CAPTCHA technique based on drag and drop mouse action. *International Journal of Technical Research and Applications*, 3(2), 188-189.
- [12] Redmon, J. & Farhadi, A. (2018). YOLOv3: An Incremental Improvement
- [13] A. S. El Ahmad, J. Yan, and M. Tayara. The robustness of google captchas. Technical report, Newcastle University, 2011.
- [14] Von Ahn, L., Maurer, B., McMillen, C., Abraham, D., & Blum, M. (2008). recaptcha: Human-based character recognition via web security measures. *Science*, 321(5895), 1465-1468.
- [15] Hussain, R., Gao, H., & Shaikh, R. A. (2017). Segmentation of connected characters in text-based CAPTCHAs for intelligent character recognition. *Multimedia Tools and Applications*, 76(24), 25547-25561.
- [16] Sharma, S., & Seth, N. (2015). Survey of Text CAPTCHA Techniques and Attacks. *International Journal of Engineering Trends and Technology (IJETT)*, 22(6).
- [17] Gao, H., Wang, W., Qi, J., Wang, X., Liu, X., & Yan, J. (2013, November). The robustness of hollow CAPTCHAs. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security* (pp. 1075-1086). ACM.
- [18] Gao, H., Tang, M., Liu, Y., Zhang, P., & Liu, X. (2017). Research on the Security of Microsoft's Two-Layer Captcha. *IEEE Transactions on Information Forensics and Security*, 12(7), 1671-1685.
- [19] Tang, M., Gao, H., Zhang, Y., Liu, Y., Zhang, P., & Wang, P. (2018). Research on Deep Learning Techniques in Breaking Text-based Captchas and Designing Image-based Captcha. *IEEE Transactions on Information Forensics and Security*, PP(99), 1556-6013.
- [20] Kumari, P., & Kapoor, M. (2015). Effect of Random Guessing Attack on Image Based Captchas: Analysis and Survey. *International Journal of Innovations & Advancement in Computer Science*, 4.
- [21] Goodfellow, I. J., Bulatov, Y., Ibarz, J., Arnoud, S., & Shet, V. Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks.
- [22] George, D., Lehrach, W., Kansky, K., Lázaro-Gredilla, M., Laan, C., Marthi, B., ... & Lavin, A. (2017). A generative vision model that trains with high data efficiency and breaks text-based CAPTCHAs. *Science*, 358(6368), eaag2612.
- [23] Chow, R., Golle, P., Jakobsson, M., Wang, L., & Wang, X. (2008, February). Making captchas clickable. In *Proceedings of the 9th workshop on Mobile computing systems and applications* (pp. 91-94). ACM.
- [24] Gatys, L. A., Ecker, A. S., & Bethge, M. (2015). A Neural Algorithm of Artistic Style. *Nature Communications*.
- [25] J. Redmon and A. Farhadi. Yolo9000: Better, faster, stronger. In *Computer Vision and Pattern Recognition (CVPR)*, 2017 IEEE Conference on, pages 6517–6525. IEEE, 2017.
- [26] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems* (pp. 91-99).
- [27] Elson, J., Douceur, J. J., Howell, J., & Saul, J. (2007). Asirra: a CAPTCHA that exploits interest-aligned manual image categorization.
- [28] Golle, P. (2008, October). Machine learning attacks against the Asirra CAPTCHA. In *Proceedings of the 15th ACM conference on Computer and communications security* (pp. 535-542). ACM.
- [29] James, A., George, G., & Yeldose, A. (2014). A Survey on Spelling Based CAPTCHA. *IJRCCCT*, 3(3), 001-007.
- [30] Kopp, M., Nikl, M., & Holena, M. Breaking CAPTCHAs with Convolutional Neural Networks.
- [31] Soumya, K.R., & Abraham, R. M. (2014). A Survey on Different CAPTCHA Techniques.
- [32] Vikram, S., Fan, Y., & Gu, G. (2011, December). SEMAGE: a new image-based two-factor CAPTCHA. In *Proceedings of the 27th Annual Computer Security Applications Conference* (pp. 237-246). ACM.
- [33] Goswami, G., Powell, B. M., Vatsa, M., Singh, R., & Noore, A. (2014). FR-CAPTCHA: CAPTCHA based on recognizing human faces. *PLoS one*, 9(4), e91708.
- [34] Goswami, G., Powell, B. M., Vatsa, M., Singh, R., & Noore, A. (2014). FaceDCAPTCHA: Face detection based color image CAPTCHA. *Future Generation Computer Systems*, 31, 59-68.
- [35] Singh, V. P., & Pal, P. (2014). Survey of different types of CAPTCHA. *International Journal of Computer Science and Information Technologies*, 5(2), 2242-2245.
- [36] Gao, H., Lei, L., Zhou, X., Li, J., & Liu, X. (2015, October). The robustness of face-based CAPTCHAs. In *Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM)*, 2015 IEEE International Conference on (pp. 2248-2255). IEEE.
- [37] Gossweiler, R., Kamvar, M., & Baluja, S. (2009, April). What's up CAPTCHA?: a CAPTCHA based on image orientation. In *Proceedings of the 18th international conference on World wide web* (pp. 841-850). ACM.
- [38] Zhao, B., Weng, H., Ji, S., Chen, J., Wang, T., He, Q., & Beyah, R. (2018, October). Towards Evaluating the Security of Real-World Deployed Image CAPTCHAs. In *Proceedings of the 11th ACM Workshop on Artificial Intelligence and Security* (pp. 85-96). ACM.
- [39] SivaKorn, S., Polakis, I., & Keromytis, A. D. (2016, March). I am robot: (deep) learning to break semantic image captchas. In *Security and Privacy (EuroS&P)*, 2016 IEEE European Symposium on (pp. 388-403). IEEE.
- [40] Girshick, Ross. "Fast R-CNN." *IEEE International Conference on Computer Vision IEEE*, 2015:1440-1448.
- [41] Gao, H., Wang, W., Fan, Y., Qi, J., & Liu, X. (2014). The Robustness of "Connecting Characters Together" CAPTCHAs. *J. Inf. Sci. Eng.*, 30(2), 347-369.
- [42] Kluever, K. A., & Zanibbi, R. (2009). Balancing usability and security in a video CAPTCHA. *Symposium on Usable Privacy and Security, SOUPS 2009, Mountain View, California, Usa, July* (pp. 1-11). DBLP.

- [43] Yan, J., & El Ahmad, A. S. (2009). Captcha security: A case study. *IEEE Security & Privacy*, 7(4).
- [44] Tam, J., Simsa, J., Hyde, S., & Ahn, L. V. (2009). Breaking audio captchas. In *Advances in Neural Information Processing Systems* (pp. 1625-1632).
- [45] Shirali-Shahreza, S., Abolhassani, H., Sameti, H., & Hassan, M. (2009, August). Spoken captcha: A captcha system for blind users. In *Electronic Commerce and Security (ISECS), 2009. CCCM 2009. ISECS International Colloquium on* (Vol. 1, pp. 221-224). IEEE.
- [46] Shirali-Shahreza, M., & Shirali-Shahreza, S. (2008, May). Motion captcha. In *Human System Interactions, 2008 Conference on* (pp. 1042-1044). IEEE.
- [47] Gao, H., Liu, H., Yao, D., Liu, X., & Aickelin, U. (2010, July). An audio CAPTCHA to distinguish humans from computers. In *Electronic Commerce and Security (ISECS), 2010 Third International Symposium on* (pp. 265-269). IEEE.
- [48] Lazar, J., Feng, J., Brooks, T., Melamed, G., Wentz, B., Holman, J., ... & Ekedebe, N. (2012, May). The SoundsRight CAPTCHA: an improved approach to audio human interaction proofs for blind users. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2267-2276). ACM.
- [49] Bursztein, E., Beauxis, R., Paskov, H., Perito, D., Fabry, C., & Mitchell, J. (2011, May). The failure of noise-based non-continuous audio captchas. In *Security and Privacy (SP), 2011 IEEE Symposium on* (pp. 19-31). IEEE.
- [50] Meutznier, H., Gupta, S., & Kolossa, D. (2015, April). Constructing secure audio captchas by exploiting differences between humans and machines. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 2335-2338). ACM.
- [51] Chen, J., Luo, X., Guo, Y., Zhang, Y., & Gong, D. (2017). A survey on breaking technique of text-based captcha. *Security & Communication Networks*, 2017(1-2), 1-15.
- [52] Sano, S., Otsuka, T., & Okuno, H. G. (2013, November). Solving Google's continuous audio CAPTCHA with HMM-based automatic speech recognition. In *International Workshop on Security* (pp. 36-52). Springer, Berlin, Heidelberg.
- [53] Dai J, Li Y, He K, et al. R-FCN: Object Detection via Region-based Fully Convolutional Networks[J]. 2016
- [54] Bursztein, E., & Bethard, S. (2009, August). Decaptcha: breaking 75% of eBay audio CAPTCHAs. In *Proceedings of the 3rd USENIX conference on Offensive technologies* (p. 8). USENIX Association.
- [55] Rao, K., Sri, K., & Sai, G. (2016). A Novel Video CAPTCHA Technique To Prevent BOT Attacks. *Procedia Computer Science*, 85, 236-240.
- [56] Converse, T. (2005). CAPTCHA generation as a web service. In *Human Interactive Proofs* (pp. 82-96). Springer, Berlin, Heidelberg.
- [57] Ya, H., Sun, H., Helt, J., & Lee, T. S. (2017). Learning to Associate Words and Images Using a Large-scale Graph. *arXiv preprint arXiv:1705.07768*.
- [58] Johnson, J., Hariharan, B., van der Maaten, L., Hoffman, J., Fei-Fei, L., Zitnick, C. L., & Girshick, R. Inferring and Executing Programs for Visual Reasoning.
- [59] D'Souza, D., Polina, P. C., & Yampolskiy, R. V. (2012). Avatar CAPTCHA: Telling computers and humans apart via face classification. *IEEE International Conference on Electro/information Technology*.
- [60] Kuppusamy, K. S., & Aghila, G. (2017). Human: an accessible, polymorphic and personalized captcha interface with preemption feature tailored for persons with visual impairments. *Universal Access in the Information Society*.
- [61] Meutznier, H., & Kolossa, D. (2016). A non-speech audio CAPTCHA based on acoustic event detection and classification. *Signal Processing Conference. IEEE*.
- [62] Moy, G., Jones, N., Harkless, C., & Potter, R. (2004, June). Distortion estimation techniques in solving visual CAPTCHAs. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on* (Vol. 2, pp. II-II). IEEE.
- [63] Liang, Z., ShuGuang, H., & Zhaoxiang, S. (2011). A highly reliable CAPTCHA recognition algorithm based on rejection. *Acta Automatica Sinica*, 37(7), 891-900.
- [64] Rui, C., Jing, Y., Rong-gui, H., & Shu-guang, H. (2013, September). A novel LSTM-RNN decoding algorithm in CAPTCHA recognition. In *Instrumentation, Measurement, Computer, Communication and Control (IMCCC), 2013 Third International Conference on* (pp. 766-771). IEEE.
- [65] Sano, S., Otsuka, T., Itoyama, K., & Okuno, H. G. (2015). HMM-based Attacks on Google's ReCAPTCHA with Continuous Visual and Audio Symbols. *Journal of Information Processing*, 23(6), 814-826.
- [66] Fiot, J. B., & Paucher, R. (2009). The captchacker project. *Ecole Centrale Paris*.
- [67] Redmon, Joseph, et al. "You Only Look Once: Unified, Real-Time Object Detection." (2015): 779-788
- [68] Yan, J., & El Ahmad, A. S. (2007, December). Breaking visual captchas with naive pattern recognition algorithms. In *Computer Security Applications Conference, 2007. ACSAC 2007. Twenty-Third Annual* (pp. 279-291). IEEE.
- [69] Gao, H., Yan, J., Cao, F., Zhang, Z., Lei, L., Tang, M., ... & Li, J. (2016). A Simple Generic Attack on Text Captchas. In *NDSS*.
- [70] Kapoor, D., Bangar, H., & Sethi, A. (2012, December). An ingenious technique for symbol identification from high noise CAPTCHA images. In *India Conference (INDICON), 2012 Annual IEEE* (pp. 098-103). IEEE.
- [71] Ye, Guixin and Tang, Zhanyong and Fang, Dingyi and Zhu, Zhanxing and Feng, Yansong and Xu, Pengfei and Chen, Xiaojiang and Wang, Zheng (2018) *Yet Another Text Captcha Solver: A Generative Adversarial Network Based Approach*. In: 25th ACM Conference on Computer and Communications Security (CCS). ACM, New York, pp. 332-348. ISBN 9781450356930
- [72] Chow, Y. W., Susilo, W., & Thorncharoensri, P. (2019). CAPTCHA Design and Security Issues. In *Advances in Cyber Security: Principles, Techniques, and Applications* (pp. 69-92). Springer, Singapore.
- [73] Chen, J., Luo, X., Liu, Y., Wang, J., & Ma, Y. (2019). Selective Learning Confusion Class for Text-Based CAPTCHA Recognition. *IEEE Access*, 7, 22246-22259.
- [74] Usmani, A., Maryam, A., Umar, M. S., & Khan, M. H. (2019). New Text-Based User Authentication Scheme Using CAPTCHA. In *Information and Communication Technology for Competitive Strategies* (pp. 313-322). Springer, Singapore.
- [75] Wu, Q. Q., Lang, J. J., Wei, S. J., Ren, M. L., & Seidel, E. (2019). A Novel Construction of Correlation-Based Image CAPTCHA with Random Walk. In *Smart Innovations in Communication and Computational Sciences* (pp. 339-346). Springer, Singapore.