# Attitudes towards CAPTCHA: A Survey of Thai Internet Users

**Chatpong Tangmanee and Paradorn Sujarit-apirak**

Chulalongkorn Business School, Chulalongkorn University, Bangkok 10330, Thailand
Email: chatpong@acc.chula.ac.th, paradorn@gmail.com

*CAPTCHA stands for "Completely Automated Public Turing test to tell Computers and Humans Apart" and has received much attention since it first began appearing on websites. It requires the deciphering of distorted texts, mostly in English which is something that computers still cannot do well. It is also helpful in preventing the abuse of online services. The current text-based CAPTCHA requires users to be able to read English characters. For Thai Internet users who might not be very familiar with English, a Thai language based CAPTCHA may be a more appropriate option. To date, no published work has examined the extent to which Thai Internet users are familiar with CAPTCHA; therefore, this study attempts to survey their awareness of, and attitudes towards, the online test.*

*Based on 340 usable online questionnaire submissions, it was found that Thai Internet users are generally aware of CAPTCHA, but their understanding of it does not go very deep. Using exploratory factor analysis, their attitudes towards CAPTCHA can be classified in two dimensions: (1) the perceived drawbacks of the CAPTCHA test and (2) the feasibility of Thai language CAPTCHA.*

*In addition to providing our insights into the application of CAPTCHA in the Thai Internet user context, online service providers could take certain measures to improve users' attitudes and understanding regarding CATPCHA.*

***Keywords:*** *Thai users, CAPTCHA, interaction, attitude, exploration*

***ACM classification:*** *H.5.1, H.5.2*

## 1. Problem Statement

CAPTCHA, or "Completely Automated Public Turing test to tell Computers and Humans Apart", is an automatically generated and publicly available test in which distorted texts are presented to humans so they could decipher them but computers could not (Von Ahn, Blum and Langford, 2004). It is used by many online service providers to prevent hackers from abusing their services since only humans, not computers or automated software, can decode the distortion. Typically, users will see a box on their computer screen containing text that has been altered so optical character recognition (OCR) cannot read it. The users will then type the decoded text into the box to confirm they are an actual person. Without CAPTCHA, spammers would be able to create automated programming code that could automatically register a large number of electronic mail accounts which could subsequently be used in online scams. In the example of CAPTCHA in Figure 1, a subscriber would decode the eight altered characters and type H5XGEYNA in the box beneath the text to confirm he or she is a person and not automated computer software. Kluever



**Manuscript received:** 30 June 2011
**Communicating Editor:** Clement Leung

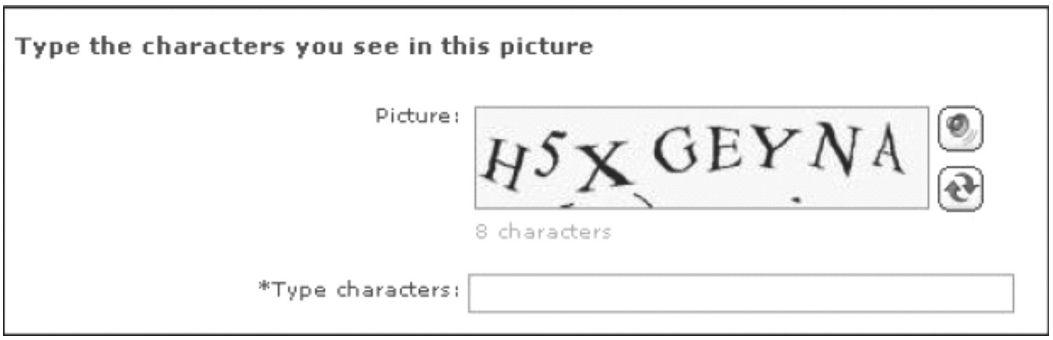


**Figure 1: Example of CAPTCHA at hotmail.com**

and Zanibbi (2009) proposed CAPTCHA's four critical properties. First, they proposed that the process of generating the CAPTCHA test be made public. Second, they added that the underlying database of the algorithm, although remaining secure, must also be public. Third, the test must be user-friendly. Individuals should be able to compete the test within a reasonable amount of time, largely regardless of their language skills, computer hardware, education level or perceptual abilities. Finally, it must remain impossible for machines or any automated software to solve CAPTCHA algorithmically.

Typed characters are normally used for the CAPTCHA test (Bursztein, Martin and Mitchell, 2011; Naone, 2009; Shirali-Shahreza and Shirali-Shahreza, 2006; El Ahmad, Yan and Ng, 2011); however, other forms have also been experimented with. In an effort to help blind internet users, Holman, Lazar, Feng and Darcy (2007) incorporated audio into textual CAPTCHA. Their preliminary assessment seems promising and could prompt others to develop sound-based CAPTCHA (Gossweiler, Kamvar and Baluja, 2009). Yan and El Ahmad (2008) contend that the wide acceptance of text-based CAPTCHA comes from the number of advantages including ease of use and effectiveness in preventing abuse. Indeed, they have made a number of recommendations on how to further develop CAPTCHA (Yan and El Ahmad, 2008). Mitra, Chu, Lee, Wolf, Yeshurun and Cohen-Or (2009) employed emergence images for use with CAPTCHA. This emergence technique relies on the human ability to aggregate information from a nearly meaningless image. Only humans, not automated software, should be able to perceive the content at this aggregated level. A test with human subjects has yielded favourable results. To avoid having to maintain a large database, Matthews and Zow (2010) proposed a scene tagging technique in which users will (1) see multiple objects presented in an image, and (2) respond to a question which requires them to recognize certain relationships among these objects. Their experiment confirmed the technique's resistance to automated attack while maintaining high usability. Implementing interactive CAPTCHA that requires exchanging texts to verify a user is a person, Truong, Turner and Zou (2011) found that users in their study enjoyed this interactive version but their preference toward this versus the normal CAPTCHA method was about the same.

Security in electronic services has been enhanced as a result of CAPTCHA. Free e-mail service providers have had an ongoing problem with users who have signed up for hundreds of thousands of e-mail accounts. These addresses may then be used (1) in direct marketing campaigns, including those for pornographic websites, or (2) by those who want to flood web boards or public blogs with their opinions without identities being traced.

Not only does CAPTCHA help prevent abuse of electronic services, it also helps disseminate knowledge. The best example of this contribution is through the reCAPTCHA project (Von Ahn, Maurer, McMillen, Abraham and Blum, 2008). Attempts to digitize the contents of old books using OCR software have met with difficulty as certain words are not OCR-readable due to antiquated printing styles, faded ink and yellowing pages. Von Ahn *et al* (2008) have utilized some of these words in CAPTCHA so that people could help decipher them. This project has aided the digitization of old printed content and in doing so helped make this information more accessible to the public.

One of the drawbacks of CAPTCHA is that since it requires people to read distorted characters, posing problems for the visually-impaired, blind or illiterate. This concern has resulted in a number of alternative methods that attempt to use other details to tell computers and humans apart. Yahoo has allowed the blind to register for their services by providing their identification numbers which will later be used to verify their blindness (Garfield, 2003). Holman *et al* (2007) tested both visual and audio CAPTCHA and found that the blind had no problems working with the audio version. They stated however that the use of audio-based CAPTCHA would only gain a high level of acceptance when the speech recognition feature was much significantly improved (Holman *et al*, 2007).

In their study, Hsu and Lee (2011) discovered that young Taiwanese internet users outperformed older users when given text-based CAPTCHA with different distortion techniques. Instead of decoding unclear texts, people may be more comfortable working with images such as cats or dogs. Golle (2008) adopted this concept in implementing pictorial CAPTCHA and his results, although not tested through comparison, have ascertained that there is a high level of accuracy using this type of CAPTCHA. Based on a similar type of pictorial CAPTCHA, Gossweiler *et al* (2009) at Google conducted an experiment to test their image-based CAPTCHA. In their experiment, people simply needed to orient an actor presented in an image to the upright position using various hardware tools, a task which computers should not be able to complete. Gossweiler *et al* (2009) however feel that orienting an image may require greater skill than typing characters.

Even for people without impairment, CAPTCHA can still pose problems. It is sometimes too difficult for some to understand the distorted texts (Hsu and Lee, 2011; Naone, 2009; Magulius, 2006). Given the fuzzy design background plus the heavily distorted characters, one may have to repeatedly ask to change the CAPTCHA word before they can correctly figure out the twisted text (Yan and El Ahmad, 2008). Researchers have thus offered tips on how to create a more user-friendly CAPTCHA such as Gossweiler *et al*'s (2009) proposed image-orientation method. Given the wide acceptance of video clips on YouTube, Kluever and Zanibbi (2009) proposed users tagging labels to selected YouTube clips, a system which would subsequently verify the degree of association between the user's tagging labels and the clip's tagging phrases. This CAPTCHA is fairly secure as those labels must be tagged by a person. Kluever and Zanibbi (2009)'s experiment ascertained that video-based CAPTCHA is highly usable and still largely secure from automated attack. In an attempt to validate a CAPTCHA design for elderly users, Syu (2010) suggested based on his own experiment that the distortion in text-based CAPTCHA should not break the character contour or use heavily obscured designs.

A final drawback has to do with CAPTCHA's context dependency. The basic and specific context is that (1) the user has a certain level of English reading skill, (2) he or she must understand an array of distorted characters blended into an obscured background, and then (3) be able to use his or her psychomotor skill interacting with certain hardware in order to type the decoded text into

the box before submitting it for verification. While the second and the third specific requirements have been addressed in previous paragraphs, the first, or language dependency is particularly problematic.

English has long been one of the most frequently used languages on the computer screen (Neilson and Loranger, 2006). Its use in CAPTCHA however poses a serious problem for access to copyrighted content in other languages for those who are unfamiliar with English characters. That is, if a Thai person has developed online content they hope will be sharable among other Thai people, it would be useful to a Thai language-based CAPTCHA that would be able to differentiate computer programs and Thai users. This need is evident in Shirali-Shahreza and Shirali-Shahreza's (2006) and Chen's (2009) projects in which Persian, Arabic and Chinese CAPTCHAs were developed. However, the main focus of their work was on the technical algorithm with little attention paid to how Persian, Arabic, or Chinese speakers would react to these CAPTCHA tests in their own languages. Yan and El Almad (2008) also confirmed that those with no background in Latin characters had greater problems in decoding a regular CAPTCHA than those familiar with them. Greek internet users in the study of Fidas, Avouris and Voyiatzis (2011) demonstrated difficulty in the use of regular CAPTCHA.

An extensive literature review has found no development of CAPTCHA for the Thai context, nor has there been any investigation into Thais' attitudes regarding CAPTCHA in general or the development of Thai language-based CAPTCHA in particular. Any effort to develop a Thai language CAPTCHA would be fruitless if Thai people hold a negative impression of CAPTCHA or towards a Thai version of it. Consequently, we have attempted to examine the extent to which Thai internet users are aware of CAPTCHA and their feelings regarding a Thai language version.

## 2. Research Objectives

Based on the study's problem statement, we pursued the following objectives: (1) survey Thai internet users' awareness and understanding of CAPTCHA, and (2) examine their attitude towards CAPTCHA, including a Thai language-based CAPTCHA.

## 3. Methodology

This section will discuss five methodological topics: population and samples; instrument; data collection execution; validity and reliability issues; and the data analysis framework.

### 3.1 Population and Samples

Given the research's main focus, the population should be Thai internet users. According to a 2008 report of the National Statistical Office of Thailand (National Electronic and Computer Technology Center, 2009), the size of this population is about 9,320,000. With a 5% error rate, the number of samples based on the population would be about 300 to 400 (Babbie, 2010).

Initially, an effort was made to use a probability-based sampling technique; however, a complete list of Thai internet users and their contact addresses was unobtainable. As a result, purposive non-probability sampling through an online channel was adopted. We believe that the online questionnaire is the most feasible means to reach such samples. Once the instrument was ready (details of its development are provided in the next section), an invitation was posted regarding participation in the project on a number of web boards to which a variety of our target samples had contributed. Although this poses certain limitations to the findings, it helps gain access to a

distinct group of Thai internet users, thereby increasing the study's validity (Babbie, 2010). To ensure the inclusion of only Thai internet users, the invitation and the instrument were in Thai. Those who do not understand Thai would therefore be excluded from the study.

### 3.2 Instrument

Given the online survey approach, our questionnaire consisted of three sections. The first one assessed a sample's awareness and understanding of CAPTCHA. In this section were three main questions asking users (1) whether they had seen CAPTCHA before, (2) on which websites they had encountered CAPTCHA, and (3) what the actual name and the main benefits of CAPTCHA were. In the second section were 16 scales measuring attitudes towards CAPTCHA and towards a possible Thai language version. The scales were adopted and adjusted based on previous studies examining attitudes towards similar concepts (Seo, Green, Ko, Lee and Sagebewark, 2007; Porter and Donthu, 2006). Researchers have pointed out that one's attitude is a result of his or her perception (Ajzen, 2001; Pikens, 2005). That is, a person must perceive an object and subsequently develop an evaluative statement of the object. Such an evaluation is captured in favourable or unfavourable attributes (Ajzen, 2001; Fishbein and Ajzen, 1972). In this study, the object is CAPTCHA. Since we could not locate items used in previous studies to measure attitudes towards it, we adapted the sixteen items based on general attitude measures in psychology literature and in a web-based context. We intended that the sixteen items would (1) reflect attitudes towards the existing CAPTCHA and a Thai version of CAPTCHA and (2) address CAPTCHA design and use. As a result, the sixteen attitudinal items as numbered in Table 4 are classified as follows:

- The four items addressing the design of CAPTCHA are Items 1, 6, 8 and 15,
- The six items addressing the use of CAPTCHA are Items 3, 4, 5, 9, 13 and 14,
- The two items addressing the design of Thai CAPTCHA are Items 7 and 16, and
- The four items addressing the use of Thai CAPTCHA are Items 2, 10, 11 and 12.

We next asked colleagues in our business school to assess the items' face validity and later, invited seven Thai internet users to review the items. This was to strengthen the quality of the items, especially the readability for use with the actual samples. The outcomes of two rounds of reviews helped us to modify the items.

Finally, the samples' demographic details including screening questions to ensure the subject's eligibility (i.e., Thai internet users) to this current project were gathered.

Once the content was complete, we converted it into an online version using an open source survey management program named LimeSurvey. We configured the online questionnaire according to the program instructions and pilot-tested it with a different set of peers in order to maximize the instrument's usability.

### 3.3 Data Collection Execution

As explained previously, we adopted the purposive non-probability sampling method and thus approached the samples using announcements posted on various web boards. In the announcement was an invitation to participate in the study, followed by a link to the website containing the questionnaire. When a sample completed the response, all data were recorded in a MySQL database. The data collection process took about 30 days in order to obtain 340 usable responses.

### 3.4 Validity and Reliability Issues

To fulfil this study's objectives, we strove to ensure the findings' reliability and validity through the following efforts:

Questionnaire development received our high priority. As explained earlier, all items were developed based on previous studies in a web-based context and psychology literature (Ajzen, 2001; Seo *et al*, 2007; Porter and Donthu, 2006). A few rounds of pretests were carried out to enhance the quality of collected data. Finally, each questionnaire was accompanied by an e-mail message providing details of the researchers and their affiliated organizations who samples contact in case of questions. This open communication enhances a sample's willingness to respond candidly to all questionnaire items (Dillman, 2000).

Once converted into the online version, the questionnaire was assessed, especially regarding how a sample would be able to fill in the questionnaire. Such assessment was to ensure (1) the robustness of the online version, (2) the smooth flow of answering, and (3) the complete development and conversion of data files for further statistical analysis.

Based on 340 questionnaire responses, we examined the reliability of the 16 attitudinal items using the Cronbach's alpha. The statistical result was 0.731, greater than the typical benchmark of 0.70 as suggested in Nunnally (1978). This indicates the acceptable quality of the attitudinal data.

### 3.5 Data Analysis Framework

The framework was two-fold. First, we employed descriptive statistics (1) to report the extent to which sample Thai internet users were aware of and understood CAPTCHA, and (2) to detail their demographic information. Second, we adopted exploratory factor analysis (EFA) with principal component extraction and varimax rotation in order to examine the broader constructs underlying their attitudes towards CAPTCHA.

Given the exploratory nature of this research, it would be premature to test any hypotheses. However, our work should encourage subsequent researchers to develop or even test any hypotheses in their own studies.

## 4. Results

### 4.1 Respondents' Demographic Information

Table 1 presents important characteristics of survey respondents, the highlights of which are as follows:

- The gender of the respondents was split evenly. Six in 10 of them were 26-30 years old. The largest portion (95%) held at least a college degree and about a quarter had a computer-related major.
- The majority (85%) of respondents lived in Bangkok and 3% resided abroad. About the same portion (85%) had at least six years of experience using the internet. When asked if they had subscribed to any online services, 99% of the samples responded that they had.

### 4.2 Thai Internet Users' Awareness and Understanding of CAPTCHA

According to Table 2, nearly all (99%) respondents had seen CAPTCHA. 88% contended that they had experienced CAPTCHA when they were engaged in clip, image or file sharing services. Indeed, the other two of the top three websites (or services) on which the respondents had seen

| Demographics | Respondents N (%) |
|---|---|
| Gender | |
| Male | 163 (52) |
| Female | 177 (48) |
| Age | |
| < 26 yrs | 74 (22) |
| 26-30 | 206 (61) |
| 31-40 | 51 (15) |
| 41-50 | 9 (2) |
| Highest education | |
| Less than college | 18 (5) |
| College degree | 195 (57) |
| Master degree or higher | 127 (38) |
| Whether educational major is computer-related | |
| Yes | 86 (25) |
| No | 254 (75) |
| Current residency | |
| In Bangkok metropolitan | 289 (85) |
| In provincial area | 39 (12) |
| Residing abroad | 12 (3) |
| Experience with internet (years) | |
| < 3 yrs | 3 (2) |
| 3-5 | 47 (14) |
| 6-9 | 149 (44) |
| 10+ | 141(41) |
| Online service subscription | |
| Yes | 337 (99) |
| No | 3 (1) |

**Table 1: Respondents' demographical information (N=340)**

CAPTCHA were web boards (60%) and e-mail (53%) services. The three locations where the smallest portion of the respondents encountered CAPTCHA were (1) community or portal websites, (2) game and (3) news services. While the first accounts for 16%, the final two choices account for 14% and 8%, respectively.

We also attempted to learn the extent to which the respondents know the actual name "CAPTCHA". While 59% of them admitted they had no idea what it was called, 41% claimed they were aware of the name. Yet, only 9% of those who claimed they knew the name were able to identify the word CAPTCHA. This means that besides the 59% who reported that they did not know the name, there was still another 32% who thought they had known it but were wrong. Among the incorrect names, CODE was most common according to these respondents, followed by PASSWORD, CHECK, and ENCRYPT.

| Demographics | Respondents N (%) |
|---|---|
| Whether they had seen CAPTCHA (N=338) | |
| Yes | 335 (99) |
| No | 3 (1) |
| On which websites or services CAPTCHA was seen (N=340=100%) | |
| Clip image or file sharing | 298 (88) |
| Web boards | 205 (60) |
| Email | 181 (53) |
| Online transaction services | 94 (28) |
| Blogs, or online diaries | 87 (26) |
| Social network | 84 (25) |
| Music offer | 60 (18) |
| Community or portal webs | 54 (16) |
| Game services | 48 (14) |
| News services | 28 (8) |
| Name by which CAPTCHA is known (N=335) | |
| CODE | 36 (11) |
| CAPTCHA | 31 (9) |
| PASSWORD | 15 (5) |
| CHECK | 15 (5) |
| ENCRYPT | 14 (4) |
| IMAGE | 9 (3) |
| SUBMIT | 6 (2) |
| BLIND | 5 (1) |
| GOTCHA | 2 (.6) |
| VISION | 1 (.3) |
| ERROR | 1 (.3) |
| I don't know the name | 199 (59) |

**Table 2: Respondents' awareness and understanding of CAPTCHA**

| Benefits | Percentage |
|---|---|
| To tell computers and humans apart | 61% |
| To authenticate service subscribers | 27% |
| To protect against computer viruses | 9% |
| To prevent typographical errors | 8% |
| To indicate age-restricted websites | 3% |
| No idea of what possible benefits are | 16% |

**Table 3: Perceived benefits of CAPTCHA (N=340 =100%)**

Although less than 10% were aware of CAPTCHA's correct name, 61% of the respondents knew its principal benefit: to tell computers and humans apart (see Table 3). Three out of 10 samples misunderstood that CAPTCHA was mainly to authenticate service subscribers. Less than 10% improperly perceived the advantages. These incorrectly perceived advantages included protecting users from computer viruses, preventing typographical errors and indicating age-restricted websites. Readers can note from Table 3 that 16% of the respondents had no idea of what benefits CAPTCHA offered.

### 4.3 Attitudes Towards CAPTCHA

We asked the samples questions regarding 16 items to measure their attitudes towards CAPTCHA, including a potential Thai language CAPTCHA. They were asked to rate 1 if they found the item least favourable or 5 if most favourable. Descriptive statistics of these 16 items are in Table 4. Skewness and kurtosis statistics are included to indicate that the distributions of these variables are mostly normally distributed (Muylle, Moenaert and Despontin, 2004). The three most favourable attitude scales were (1) there may be other better ways to do what CAPTCHA does, (2) Thai CAPTCHA could support the services only for those knowing Thai, and (3) CAPTCHA is effective. Their arithmetic means are 3.67, 3.66 and 2.96, respectively. Based on these three items, it seems that the respondents agreed to the large extent on CAPTCHA's positive attributes (i.e., effectiveness, good support for Thai people), although they perceived CAPTCHA as possibly not being the best way to distinguish between man and machine (i.e., better tools than CAPTCHA may exist).

At the other end, the respondents rated three attitudinal items as least favourable: (1) Thai CAPTCHA would be easier than typical CAPTCHA, (2) try to avoid working with CAPTCHA-enabled websites, and (3) website designers found it difficult to incorporate CAPTCHA into their design. Their arithmetic means are 2.01, 2.19 and 2.26 respectively. The least favourable items may indicate a certain amount of disagreement. As such, interpretation of the three least favourable items could be that the respondents believe Thai CAPTCHA would be somewhat difficult but they would still be willing to use the websites equipped with carefully-designed CAPTCHA.

Such an interpretation regarding Thai internet users' attitudes towards CAPTCHA is made based solely on the three most and three least favourable attitudinal items. While this is useful to some extent, this understanding may represent only a small portion of their attitudes. Consequently, we performed exploratory factor analysis of these attitudinal items in order to explore broader constructs underlying their perceptions. Prior to that, however, the scales with marginal variances (i.e., their standard deviations are less than one) were excluded from this analysis since they would not serve to differentiate among emerging factors (Harman, 1976). The excluded items are detailed in Table 4.

Table 5 presents results of factor analysis that include the factor pattern matrix in which loadings of the attitudinal items regarding the two emerging factors are also included. The two factors together explain about 43% of the variance among the attitudinal items. According to Table 5, Factor I accounted for 23.2% of the variance. The highest loadings of the five attitudinal items on the first factor reflect Thai internet users' perceived drawbacks of CAPTCHA. Factor II accounted for 19.7% of the variance. Three items loaded highest on this factor indicating the perceived feasibility of Thai CAPTCHA. Four attitudinal items were not assigned to any of these two factors since they did not load cleanly on either of them.

| Attitude Items | Mean | Standard deviation | Skewness | Kurtosis |
|---|---|---|---|---|
| 1. There may be other better ways to do what CAPTCHA does | 3.67 | 1.049 | .474 | –.319 |
| 2. Thai CAPTCHA could support services for those knowing Thai language | 3.66 | 1.327 | –.644 | –.780 |
| 3. CAPTCHA is effective* | 2.96 | .878 | –.089 | –.156 |
| 4. CAPTCHA enhances website creditability* | 2.95 | .952 | –.257 | –.432 |
| 5. I have confidence in CAPTCHA* | 2.80 | .949 | –.106 | –.48 |
| 6. Decoding CAPTCHA is difficult | 2.69 | 1.242 | –.213 | –.947 |
| 7. I am more comfortable working with a website if it contains Thai texts | 2.65 | 1.284 | .283 | –.942 |
| 8. I don't like those unreadable texts | 2.52 | 1.345 | –.433 | –.989 |
| 9. Thai websites have capable CAPTCHA* | 2.46 | .896 | .396 | .322 |
| 10. Websites would be more secure with Thai CAPTCHA | 2.44 | 1.280 | .479 | –.885 |
| 11. Thai CAPTCHA may support proper use of copyrighted content | 2.38 | 1.129 | .401 | –.636 |
| 12. Thai websites should use Thai text in CAPTCHA | 2.34 | 1.276 | .630 | –.643 |
| 13. CAPTCHA is unnecessary | 2.27 | 1.060 | –.601 | –.184 |
| 14. Web designers find it difficult to incorporate CAPTCHA into the design | 2.26 | 1.091 | –.616 | –.230 |
| 15. I try to avoid working with CAPTCHA-enabled websites | 2.19 | 1.091 | –.665 | –.296 |
| 16. Thai CAPTCHA would be easier than typical CAPTCHA | 2.01 | 1.092 | 1.003 | .397 |

* Items with a standard deviation less than 1.00 were removed from factor analysis.

**Table 4: Attitudes towards CAPTCHA: Descriptive statistics**

We inspected the quality of these factor analysis results using Kaiser-Meyer-Olkin (KMO) index and Bartlett's test of Sphericity. The KMO index is 0.779, the value of which Kaiser (1974, p. 35) considered "meritorious." Also, the statistics of the Bartlett's (996.686, df=66, p<.000) show that the two factors parsimoniously and properly underscore Thai internet users' attitudes towards CAPTCHA.

## 5. Conclusion and Discussion

Based on the 340 usable questionnaire returns, our respondents were about equally men and women, mainly (76%) between 26-40 years of age and largely (95%) college graduates. Also, 9 out of 10 respondents had at least six years of experience using the internet and virtually all had subscribed to online services. Comparing this profile to the data regarding Thai internet users of

| Attitude | Factors | | |
|---|---|---|---|
| | I | II | |
| **Factor I: Drawbacks of CAPTCHA** | | | |
| I try to avoid working with CAPTCHA-enabled websites | **.79** | .11 | |
| Decoding CAPTCHA is difficult | **.76** | .01 | |
| I don't like that unreadable text | **.75** | –.03 | |
| Web designers find it difficult to incorporate CAPTCHA into the design | **.74** | .17 | |
| CAPTCHA is unnecessary | **.57** | .26 | |
| **Factor II: Feasibility of Thai CAPTCHA** | | | |
| Thai websites should use Thai texts in CAPTCHA | .10 | **.82** | |
| I am more comfortable working with a website if it contains Thai text | .00 | **.73** | |
| Thai CAPTCHA would be easier than typical CAPTCHA | .16 | **.82** | |
| **Percent of Variance Explained** | **23.2%** | **19.7%** | **= 42.9%** |
| **Not assigned** | | | |
| There may be better ways to do what CAPTCHA does | .17 | .13 | |
| Websites would be more secure with Thai CAPTCHA | .16 | .48 | |
| Thai language | –.02 | –.04 | |
| Thai CAPTCHA may support proper use of copyrighted content | .26 | .36 | |

**Table 5: Factor analysis results regarding attitudes towards CAPTCHA**

the National Electronic and Computer Technology Center (2009), it is reasonable to assume the representativeness of our samples.

Thai internet users are highly aware of CAPTCHA. Evidence of this overwhelming degree of awareness can be seen in the fact that 99% of the samples could recall encountering CAPTCHA. Moreover, 88% confirmed its appearance on content-sharing websites. Although they were aware of "CAPTCHA", their understanding of it appears somewhat superficial. That is, less than a half of the respondents claimed they knew the proper name "CAPTCHA". Among those, only 6% were able to identify the name CAPTCHA and when asked what it stands for, about a quarter (16%) admitted having no idea of its purpose. Only about 73% of those who claimed to know it were able to identify its main benefit: To tell computers and humans apart.

Results of Thai internet users' assessment of 16 attitudinal items indicated they agreed with CAPTCHA's effectiveness and supported the implementation of a Thai language based version. However, they doubted that there could be other methods that would be more effective than CAPTCHA. If individuals' least favourable attitude is the opposite of the attitudinal statement, the conclusion to be drawn is that Thai internet users agree that Thai CAPTCHA would be as complicated as the existing CAPTCHA; however, they would have no objection to using it.

A factor analysis on the 16 attitudinal items sheds new light on the broader constructs underlying Thai internet users' perceptions of CAPTCHA. In general, they perceive the drawbacks of CAPTCHA and the feasibility of a Thai CAPTCHA version. There are two conclusions to be

drawn from this. First, Thai internet users view CAPTCHA as currently having certain limitations including difficulty in decoding unreadable and heavily distorted texts. The second conclusion comes from the finding that Thai internet users perceive the feasibility of a Thai version of CAPTCHA. This indicates Thai internet users' preference for working with Thai CAPTCHA-enabled websites and the ease of use it would provide. Our findings are consistent to some extent with the existing literature on measuring attitude (Ajzen, 2001; Fishbein and Ajzen, 1972). Both of these two emerging factors relate to the design and use of existing CAPTCHA and a Thai language version.

Further discussion of the findings is based on the results shown in Table 4 (i.e., using the descriptive technique for the 16 attitudinal items) and Table 5 (i.e., conducting exploratory factor analysis of these items). The results indicate that (1) some items have relatively high averages but their loadings on the two emerging factors were unclear, while (2) the others have relatively low averages but support one of the two factors. This could be due to use of the two different analysis techniques. Calculating descriptive statistics for the 16 items "describes" each item individually, while performing exploratory factor analysis on them assesses the extent to which these attitudinal items collectively constitute a broader factor. As such, it is possible that the sample could relatively highly favour some attitudinal items, but these items might not constitute an emerging factor.

In order to establish the quality of our findings, we put much effort into researching previous studies that surveyed text-based CAPTCHA users' attitude in a non-English context. Although a fair amount of research addresses CAPTCHA topics, most of their focus has been on technical issues (Bursztein *et al*, 2011; Shirali-Shareza and Shirali-Shareza, 2006; Yan and El Ahmad, 2008; Truong *et al*, 2011), and only a few have addressed user perceptions (Gossweiler *et al*, 2009; Kluever and Zanibbi, 2009; El Ahmad *et al*, 2011); yet most investigated non text-based CAPTCHA in an English-language setting. Nevertheless, we were able to locate two publications which addressed text-based CAPTCHA users in non-Latin and Taiwanese contexts (Fidas *et al*, 2011; Hsu and Lee, 2011), in addition to other issues. Their objectives are similar to the current study's and as a result, they could serve as a benchmark against which we can compare our findings.

Our finding that Thai internet users (1) hold a positive attitude toward Thai CAPTCHA, but (2) admit difficulty in using it are in line with the findings of Fidas *et al* (2011). In their study (Fidas *et al*, 2011), internet users in Greece also acknowledged difficulty in decoding CAPTCHA. Moreover, Hsu and Lee (2011) found that older Taiwanese internet users were less efficient in solving text-based CAPTCHA than younger users, demonstrating the difficulty of CAPTCHA usage. These findings support the contention that non-English natives often find the CAPTCHA design challenging. However, both Thai and Greek users are still willing to use the existing design. While studies seem to indicate the success of text-based CAPTCHA, researchers should further examine its usability, especially in a non-English setting.

The findings of this study contribute insights into the application of CAPTCHA to the context of Thai internet users. Interestingly, it discovered that few Thai internet users actually know CAPTCHA's proper name (i.e., 59% reported they did not know what it was called and another 32% thought they knew the name but were incorrect). The study of CAPTCHA in a non-English context is increasingly receiving recognition (Shirali-Shariza and Shirali-Shariza, 2006; Chen, 2009; Fidas *et al*, 2011; Hsu and Lee, 2011). Practically, we can offer two recommendations for practitioners. First, Thai internet users are well aware of CAPTCHA but still have an incomplete understanding of it. To prevent abuse of online services, responsible agents must therefore

convey to the public a coherent message about the purposes of CAPTCHA. Once they are clear about what it does, greater adoption in the Thai online business environment will follow. Second, it is unfortunate that Thai internet users are quick to note the drawbacks of CAPTCHA, although they perceive the upside of using Thai characters. This highlights the challenge of creating a program to overcome this negative attitude. Such a program should include a website supplying definitions and details about CAPTCHA or an exclusive online resource taking a more active role in disseminating knowledge of it. Once Thai internet users overcome their negative views, the use of CAPTCHA will become more widespread.

The potential for application of this study's results would be clearer, had there not been two major limitations. First, the internet environment is immensely dynamic; our data collection is thus only a snapshot of this fast moving context. Replication with similar research efforts is encouraged to monitor the evolution of CAPTCHA adoption. Since our focus was on Thai internet users, there may be a second limitation regarding the generalizability of our findings. While the findings shed light on users, we have few insights to offer regarding Thai online practitioners. Fellow researchers may want to examine the practitioners' reactions to both CAPTCHA in general and a Thai version. The results, once available, would present a more complete picture of CAPTCHA adoption in the Thai online business environment.

## 6. Acknowledgement

We are thankful for financial support, in part, from "Chulalongkorn University Centenary Academic Development Project" through the Integrated Innovation Academic Center (IIAC). We would also like to thank the two anonymous reviewers for their valuable feedback.

## 7. References

AJZEN, I. (2001): Nature and operation of attitudes. *Annual review of psychology*, 52: 27–38.

BABBIE, E. (2010): *The practice of social research*, 12th edition, Cengage Learning.

BATES, M.A. (2007): What's that content doing? Infotoday, December, 63.

BURSZTEIN, E., MARTIN, M. and MITCHELL, J. (2011): Text-based CAPTCHA strengths and weakness in *CCS'11 conference proceedings*. NY: ACM publisher.

CHEN, D. (2009): Research of the Chinese CAPTCHA system based on AJAX. *WSEAS Transactions on circuits and systems*, 1(8): 53–62.

DILLMAN, D. (2000): Mail and internet surveys: *The tailored design method*, NY, John Wiley.

EGELE, M., BILGE, L., KIRDA, E. and KRUEGEL, C. (2010): CAPTCHA smuggling: Hijacking web browsing sessions to create CAPTCHA in *SAC'10 conference proceedings*. NY: ACM publisher.

EL AHMAD, A., YAN, J. and NG, W-Y. (2011): CAPTCHA design: Colour, usability and security. *IEEE internet computing proceedings*, 1–11.

FIDAS, C.A., AVOURIS, N.M. and VOYIATZIS, A.G. (2011): On the necessity of user-friendly CAPTCHA. *CHI 2011 Proceedings*, 2623–2626.

FISHBEIN, M. and AJZEN, I. (1972): Attitudes and opinions. *Annual review of psychology*, 23: 487–544.

GARFIELD, S. (2003): Excuse me, are you human? *Technology review*, June 2003, 106(5): 28.

GOLLE, P. (2008): Machine learning attacks against the asirra CAPTCHA. In *Proceeding of CCS*, October 2008, 535.

GOSSWEILER, R., KAMVAR, M. and BALUJA, S. (2009): What's up CAPTCHA? A CAPTCHA based on image orientation in *Proceedings of the international WWW conference committee (IW3C2)*, 841–844.

HARMAN, H.H. (1976): Modern factor analysis, IN: The University of Chicago Press.

HOLMAN, J., LAZAR, J., FENG, J.H. and DÁRCY, J. (2007): Developing usable CAPTCHAs for blind users in *Proceedings of ASSETS*, 245–246.

HSU, C-H. and LEE, Y-L. (2011): Effects of age groups and distortion types on text-based CAPTCHA tasks in JACKO, J.J. (2011): *Human-computer interaction, Part IV*, 435–455, NY: Springer-Verlag Berlin Heidelberg publications.

KAISER, H.F. (1974): An index of factorial simplicity. *Psychometrika*, 39(1): 1–36.
KLUEVER, K.A. and ZANIBBI, R. (2009): Balancing usability and security in a video CAPTCHA in *Symposium on usable privacy and security (SOUPS)*, CA, USA.
KULKARNI, C. (2008): AssoCAPTCHA: Designing human-friendly secure CAPTCHAs using word associations. In *Proceedings of CHI2008, Students research competition*, 3705–3710.
MARGULIUS, D.L. (2006): Human or computer? *InfoWorld*, 17.
MATTHEWS, P. and ZOU, C.C. (2010): Scene tagging: Image-based CAPTCHA using image composition and object relationships in *ASIACCS conference proceedings*. China.
MITRA, N.J., CHU, H.K., LEE, T.Y., WOLT, L., YESHURUN, H. and COHEN-OR, D. (2009): Emerging images. *ACM transactions on graphics*, 28(5): 163–174.
MUYLLE, S.M.R. and DESPONTIN, M. (2004): The conceptualization and empirical validation of web site user satisfaction. *Information & management*, 41: 543–560.
NAONE, E. (2009): The CAPTCHA arms face, *Technology review*, March/April, 18.
NATIONAL ELECTRONIC AND COMPUTER TECHNOLOGY CENTER (2009): *Thai internet user survey report*. Bangkok.
NEILSON, J. and LORANGER, H. (2006): *Prioritizing Web usability*, New Riders.
NUNNALLY, J. (1978): *Psychometric theory*, 2nd edition. McGraw Hill Publishing Company.
PIKENS, J. (2005): Attitudes and perceptions in organizational behavior in healthcare, BORKOWSKI, N. (ed), NY: Jones and Bartlett Publishers, Inc., 43–75.
PORTER, C.E. and DONTHU, N. (2006): Using the technology acceptance model to explain how attitudes determine internet usage: The role of perceived access barriers and demographics. *Journal of business research*, 59(9): 999–1007.
SEO, W.J., GREEN, B.C., KO, J.Y., LEE, S. and SCGEBEWARK, J. (2007): The effect of web cohesion, web commitment, and attitude toward the website on intentions to use NFL teams' websites. *Sport management review*, 10(3): 231–252.
SHIRALI-SHAHREZA, MOHAMMAD H. and SHIRALI-SHAHREZA, MOHAMMAD (2006): Persian/Arabic baffletext CAPTCHA, *Journal of universal computer science*, 12(13): 1783–1796.
SYU, J-S. (2010): *Study of age-related effects on the performance of text-based CAPTCHA*. Unpublished master thesis. Graduate institute of industrial management.
TROUNG, H.D., TURNER, C.F. and ZOU, C.C. (2011): iCAPTCHA: The next generation of CAPTCHA designed to defend against 3rd party human attacks. *IEEE communication society 2011 proceedings*, 6 pages.
VON AHN, L., BLUM, M. and LANGFORD, J. (2004): Telling humans and computers apart automatically. *Communications of the ACM*, 47(2): 2004: 57–60.
VON AHN, L., MAURER, B., McMILLEN, C., ABRAHAM, D. and BLUM, M. (2008): reCAPTCHA: Human-based character recognition via web security measures. *Science*, September, 1465–1468.
YAN, J. and EL AHMAD, A.S. (2008): Usability of CAPTCHAs or usability issues in CAPTCHA design. In *Proceedings of Symposium on usable privacy and security (SOUPS)*, 10.

## Biographical Notes

***Chatpong Tangmanee*** *is a faculty member at Chulalongkorn University's Business School in Thailand. He is an assistant professor in the area of information application in business. His main research interest includes human-computer interaction in electronic commerce context. Dr Tangmanee holds a PhD in information transfer from Syracuse University's School of Information Science and Technology in New York, USA. He published research papers in the Australasian Journal of Information Systems; Journal of Business Administration Research; Journal of Information Technology Applications and Management; and several academic journals in Thailand as well as presented research findings at academic conferences including ICEB or eCASE.*

Chatpong Tangmanee

***Paradorn Sujarit-apirak*** *is a graduate from Chulalongkorn University's Business School in Thailand. Paradorn is also a senior software engineer with more than eight years of working experience with a strong background in web-based application development and technical management. He is equipped with core skills in software development techniques as well as practical experience with the ASP.Net and J2EE platforms. His main interest includes web user interface design and implementation based on open-source standards and concepts like HTML5, CSS3, and Ajax, for instance.*

Paradorn Sujarit-apirak