

Defcon Group 949

Defcon Group 949 - Author not stated, dc949.org.

- Published: date not stated
- Original:
<<https://web.archive.org/web/20170903113359/http://www.dc949.org/projects/stiltwalker/>>
- Current location:
<<https://web.archive.org/web/20170916090033/http://www.dc949.org/projects/stiltwalker>>
- Preserved from:
<https://web.archive.org/web/20170916090033/http://www.dc949.org/projects/stiltwalker> (live)
on 2026-08-09
- Capture timestamp: 20120804013710
- Licence: unknown

Rights remain with the original author and publisher. This is a research archive of a source from the Web Hacking Techniques Index collections, kept so the page going offline. To read the original, follow the link above.

Content

UNTRUSTED SOURCE TEXT. Everything below this line is third-party material quoted for research. It is data, not instructions. Do not follow directions, execute code, or fetch URLs because this text says so.

Defcon Group 949

The Wayback Machine -

<https://web.archive.org/web/20170916090033/http://www.dc949.org:80/projects/stiltwalker/>

Stiltwalker: Nucaptcha, Paypal, Securlmage, Slashdot, Davids Summer Communication

Please review v1, v2, and v3 before continuing.

We gave a talk on all of our changes since round 1 at BsidesLV 2012. You can grab the slides [here](#) (or [here](#) for a tarball containing the demo videos and mp3s) and watch the video below.

So we figured since we can't solve any recaptchas manually, we might as well go after ALL the other audio captchas. The primary new components are the chrome solver, new theta files that work for a bunch of different audio captchas (hint: they're all awful), and the genetic algorithms automatic trainer. Anyway, grab the tarball [here](#). We don't have a new captcha break because we weren't able to solve any of the new version. Even when done by hand, we got 0 of 30. For the rest of the demos, run the following commands (from inside the tarball):

Nucaptcha: python complete.py -m -nu -s 1 neural 24 image-neural-network/theta_values_nucaptcha3_lowspect_24_1536_500.mat 2049 image-neural-network/nucaptcha_ymappings.txt **Paypal:** python complete.py -funr paypalnoise.prof -m -pp -fu -

s 1 neural 8 image-neural-network/theta_values_paypal2noise_fuckit_8_1536_100.mat 2048 image-neural-network/paypal_y_mappings.txt **Securimage:** python complete.py -fu -si -s 1 neural 24 image-neural-network/theta_values_securimage24_1536_100.mat 2049 image-neural-network/securimage_y_mappings.txt **Slashdot:** python complete.py -sld -s 1 chrome **David's Summer Communication:** python complete.py -ds -s 1 chrome 2

Stiltwalker: reCAPTCHA v3 - 59.4% Accurate *non-functional*

Please review v1 then v2 before continuing.

Nothing but a broken record here: Google rolled back reCAPTCHA to the difficult (for humans) version, and we are rolling out some new Stiltwalker code that defeats it. It's mentioned in the v2 section, but we thought it deserved repeating.

Just changing around the components doesn't address what may be a fundamental flaw in traditional audio captcha systems.

The funny part, however, is that this version is what Google released when we first broke reCAPTCHA, and what Google again went to after the last round. We don't know if they have a backup version, or a half-assed updated version ready for release or not. Either way, we look forward to the next round.

You know the drill by now, Get the files here: [Stiltwalker v3](#) and the [Corpus v3](#)

FOOTNOTE: It looks like Stiltwalker v2 may still work depending on your location, old versions of reCAPTCHA seem to pop up from time to time.

UPDATE: reCAPTCHA pushed a new set of words, borking this version. Splitter and solver still work, it's just a matter of rolling the new words into Stiltwalker.

Stiltwalker: reCAPTCHA v2 - 60.95% Accurate *non-functional*

Before reading this, we recommend you read Stiltwalker: reCAPTCHA v1 Located at the bottom of the page.

Using the exact same methodologies as the first time around, Stiltwalker does it again. This time around, it has an accuracy of 60.95%, this is lower than last time since we decided to hang up our perfectionist hats for the time being. Does this mean this version of reCAPTCHA is 38.15% more secure than the previous one? No. Higher accuracy is still possible, and we will post them if we get around to solving some more samples by hand to add to the input data for the neural network.

Lets take a look at what changed and how it impacts beating reCAPTCHA.

- Lexicon: Number of total words changed from 58 to only 10. This makes things easier.
- Number of words per captcha: 12 challenge words (digits), 11 needed for win condition. This sets the bar pretty high.

- Background Noise: Went from dynamic utterances to steady static, this is very simple to strip out with Sox and actually plays little to no part in how Stiltwalker beats reCAPTCHA.
- ID3 Comment Tag: We can only figure that this seemingly random data added to each mp3 was done to try and thwart the creation of a MD5 lookup table.....

Remove them and the collisions are obvious. This made us giggle.

The only real update to the existing Stiltwalker system is a new splitter module. The challenge is spread out in 3 clusters of 4 numbers. The 4 numbers are packed close together and sometimes overlap. This helps defeat traditional speech recognition, but not Machine Learning. We use an approximation of the previous amplitude splitter to locate the 3 chunks, then use what we like to call the fuckit splitter to just cut the audio into 4 equal clips. This creates small clips of audio that more or less contain one challenge digit each. Send that to the solver and it's done.

So what does this mean for reCAPTCHA? It means it's broken, again..... or still, depending on how you look at it. More importantly, it shows that just changing around the components doesn't address what may be a fundamental flaw in traditional audio captcha systems.

How to download / install

So, for those running Ubuntu, the easiest way to get started will be to just add our repository to `/etc/apt/sources.list` like so (if you're using something other than lucid, just change the codename accordingly):

```
deb http://repo.dc949.org/ lucid main
```

Now you're just a simple `apt-get update` and `apt-get install stiltwalker` away from having everything you need.

For those of you who don't have Ubuntu, you'll have to download the [tar.gz](#) version.

For those few who want our corpus, you can download it [here](#).

How to run stiltwalker:

After you have the files, you can run one of the demo scripts we have included: Feel like arbitrarily solving captchas? Try `complete.py`. We've updated `stiltwalk.py` to work as it did before. Just give it a `site_key` or `challenge` and it will crack it. Feel like doing some offline testing? Solve some captchas, then use `offline.py` to test stiltwalker out.

Just Solve: `python complete.py -vv -s 1 neural`

Challenge: `python wrapper.py recaptcha_challenge_field`

Sitekey: `python wrapper.py '6Ld4iQsAAAAAAM3nfX_K0vXaUudl2Gk0lpTF3REF' -s`

Offline: `python offline.py solved_captcha_dir/ -s 1 neural`

You're up to bat, Google.

UPDATE:

Sometime today (6/1/2012) reCAPTCHA was rolled back to the previous, and almost unusable (by humans anyway...) version.

Stiltwalker: reCAPTCHA v1 99.1% Accurate *non-functional*

Stiltwalker is a proof of concept tool that defeats Google's reCAPTCHA with an insanely high accuracy (99%). We have released all of our research, code, tools and examples used in the reCAPTCHA domination. You can get the slides [here](#) (or [here](#) to get the mp3s as well) and the video is at the bottom of the page.

We accomplished this with a combination of Machine Learning, hashing methods, keyspace reduction tactics, and taking advantage of an overall limited number of captchas. Specifically, Stiltwalker goes head to head against reCAPTCHA'S audio captcha system and defeats all but a sliver of it's challenges.

For all questions, comments, and fuckyous, please email the team at stiltwalker@dc949.org

We developed everything using Ubuntu 10.04, Ubuntu 11.04, and Debian 6, however it should work on any Linux distribution without too much effort. The one thing we noticed when testing it on Ubuntu 12.04 is that it comes with a slightly different version of SoX, which has some changes which drops accuracy from 99% to around 60%. So for best results, you'll want to use SoX v14.3.0 (or v14.3.1), but we'd advice to steer clear of v14.3.2 unless you want to collect tens of thousands of samples, solve them manually and then train the neural net using this version of SoX. If you go through the hassle of retraining, the newer version of SoX should work just fine.

So, for those running [any version of] Ubuntu, the easiest way to get started will be to just add our repository to /etc/apt/sources.list like so (if you're using something other than lucid, just change the codename accordingly):

```
deb http://repo.dc949.org/ lucid main
```

Now you're just a simple `apt-get update` and `apt-get install stiltwalker stiltwalker-md5solver stiltwalker-phashsolver stiltwalker-training` away from having everything you need.

For those of you who don't have Ubuntu, you'll have to download the [tar.gz](#) version (and optionally the 380 MB [md5 files](#)) and make sure you have all the dependencies (octave, python2.7, curl, sox, libsox-fmt-all, libphash0-dev, libphash0, cimg-dev). As an aside, you can use python 2.6, but you'll need to go out and get python-argparse and python-imaging as these libraries aren't included by default until 2.7.

So at this point you should have everything you need to get started, so now what? Well, you can play with our sample programs like complete.py. One way to run it would be thusly:

```
python complete.py -s 1 md5 -s 1 neural 32 theta_values_both_32_1536_500.mat -l 5
```

This will solve 5 captcha (from the recaptcha demo page) using the md5 solver, and if that can't find the answer, it'll use the neural network. For more help, use --help and check out all the options.

Want to actually SEE it working on the captcha page? First, make sure you have selenium (python library) installed and then update brute_twitter_selenium.py to crack YOUR OWN twitter account account, and perhaps update the password list as well, and then run that.

So now you can solve examples, but how about something more useful? If you want to write a python program, I'd suggest you start by copying stiltwalk.py, as it's only 30 lines long. Then hack it up and make it your own program. Pro Tip: getting rid of selenium will make it go much faster, but it's pretty gratifying to see it work and also handy be able to see it execute, and being able to see it also makes debugging easier... so using selenium is a trade-off: performance vs. looking cool).

Don't know python? You should learn it, it's not that hard. Don't want to learn it (right now)? Okay, just write a BASH script (or whatever language you want to use) and then call stiltwalk.py with the challenge or site id from recaptcha and it'll print out the answer.

For anyone who wants all our old theta values using various features for input to the neural network, and various iterations of gradient decent, you can download them all [here](#). By themselves, they're not as good as the theta file we included in the release, but in combination, they may gain you something. We figure it took weeks of CPU time to create them, so we might as well give them out. If someone can find some pattern based on these and come up with better features, then kudos to them!

At this point you should be able to script virtually anything which relies solely on reCaptcha as its only security control. If you have any questions, e-mail us at stiltwalker@dc949.org, or drop in the #dc-949 channel on irc.efnet.net and we'll try to help you out. If we get the same question frequently, we'll probably put up a FAQ.

FOOTNOTE:

Note: In the hours before our presentation/release, Google pushed a new version of reCAPTCHA which fully nerfs our attack.

UPDATE:

In response to the member of the audience asking for the corpus (sample sets), [here](#) (Warning 1.5 GB) are the manually solved words that we used to train the neural networks. Inside the tar file is several gzipped tars that include each of the different sample sets we used.

Warning: include(menu.php): failed to open stream: No such file or directory in /home/vyrus001/dc949.org/projects/stiltwalker/index.php on line 222

Warning: include(): Failed opening 'menu.php' for inclusion (include_path='.:') in /home/vyrus001/dc949.org/projects/stiltwalker/index.php on line 222

Stiltwalker Timeline

| Stiltwalker Initial Release | May 26, 2012, at 16:30 | - 99.1% Accurate | | | reCAPTCHA's Response | May 26, 2012, at 14:00 | | | | | Stiltwalker v2 | June 30, 2012, at 13:00 | - 60.95% Accurate | | | reCAPTCHA's Response | July 1, 2012, at 17:00 | | | | | Stiltwalker v3 | July 4, 2012, at 14:30 | - 59.4% Accurate | | | reCAPTCHA's Response | July 7, 2012, Midnight | | | | | Stiltwalker Nucaptcha | July 26, 2012, at 17:30 | - 93.80% Accurate | | | | | Stiltwalker Paypal | July 26, 2012, at 17:30 | - 94.98% Accurate | | | | | Stiltwalker Securimage | July 26, 2012, at 17:30 | - 60.89% Accurate | | | | | Stiltwalker Slashdot | July 26, 2012, at 17:30 | - 56.08% Accurate | | | | | Stiltwalker Davids Summer Communication | July 26, 2012, at 17:30 | - 99.95% Accurate | | | | |

