New records for Playfair solutions

Richard Bean
School of Electrical Engineering
and Computer Science
University of Queensland
Australia
r.bean1@uq.edu.au

Louie Helm
RockstarResearch.com
louiehelm@protonmail.ch

Abstract

We give solutions to the 24 letter and 22 letter Playfair challenges proposed in Dunin et al. (2022) A number of methods were tried combining successful approaches of previous solvers, introducing new ideas while using letter-level and word-level approaches.

We used vanilla and positional n-gram models for n values from 6 up to 10.

However, these did not greatly assist to distinguish the intended solution from other high-scoring solutions. The most effective discriminative approach involved using a multi-terabyte-scale, unpruned large language model from Buck, Heafield, and Van Ooyen (2014) which moved the solution in each case into the top 5,000 ranked possibilities.

1 Introduction

Beginning in 2019, Schmeh proposed Playfair cipher challenges on his classic cipher blog, Klausi's Krypto Kolumne (Schmeh (2019)). Previous authors such as Deavours (1977) had examined the concept of the "unicity distance" for various ciphers.

Deavours stated: "The *unicity point* of a cipher is the message length beyond which decipherment using a known system becomes a unique process. For messages shorter than the unicity point distance, plural decipherments are the rule and the would-be cryptanalyst has no possible method of selecting the correct decipherment from the many available ones."

He estimated the unicity distance of the Playfair cipher in English as 22.69 letters.

The first challenge cipher was 50 letters of ciphertext. The blog post title was "Playfair cipher: Is it unbreakable, if the message has only 50 letters?". On the same day of the post, George Lasry posted the solution.

The next 40 letter challenge was posted and solved on 8 December 2018 by Nils Kopal (described at Lasry (2019)).

Next, on 15 April 2019, a 30 letter challenge was posted. Magnus Ekhall posted the solution on 3 September 2019.

On 10 September 2019, a 28 letter challenge was posted which Magnus Ekhall solved on 11 November 2019.

Finally, Konstantin Hamidullin solved a 26 letter challenge (22 November 2019) on 17 December 2019.

Most recently, Schmeh posted a 24 letter challenge on 27 January 2020. On 13 August 2021, in a paper explaining previous challenges Dunin et al. (2022) proposed another 22 letter challenge. This length was below Deavours' estimated unicity distance.

Despite previous solvers posting potential solutions, no definitive solution was found.

2 Method

2.1 Cryptanalytic Workflow

We noted that the previous solutions had not used n-gram statistics with $n > 6$. For example, the paper Bean (2020) gave examples of two classical cipher challenges solved in 2019 through the use of 6-gram English statistics. It was expected that 8-gram statistics, in particular, would provide much greater solving power in the case of Playfair ciphers.

Previous successful solver techniques had often been based on the simulated annealing approach published by Cowan (2008) (example C++ code at Cowan (2009) and C code at Lyons (2012). This approach is also used in CrypTool Kopal (2018) with C# code.

Proceedings of the 8th International Conference on Historical Cryptology, pages 12-17
Poznań, Poland, 16-18 June, 2025

Word Count	Score	Text
4	460	WAIT FOR FURTHER INSTRUCTIONS
6	763	STAY WHERE YOU ARE UNTIL THURSDAY
8	1106	TAKE THE LAST TRAIN TO YORK ON SUNDAY
9	1061	MEET YOU TOMORROW AT FOUR TWENTY AT MARKET PLACE
11	1491	WHILE IN PARIS I RECEIVED ORDERS TO REPORT TO GENERAL FOSTER

Table 1: Initial 2-word-gram scores of previous challenges.

We combined previous techniques (simulated annealing based on Cowan (2008), plus tabu search to avoid examining past solutions) with 8-gram statistics. We used C code and "libpthread" for multithreaded execution with up to 96 threads. With this approach, we were able to solve all previous challenges relatively quickly.

This is despite the observation of Lasry (2018) that higher n values are a double-edged sword: more selective, but less resilient to key errors.

We also tested scoring with 2-word-grams, with an initial dictionary of 20,000 common English words. We found the maximum score possible by splitting up each solution into known English words using dynamic programming. Each pair of words was scored by frequency, logged and then scaled to one byte. Table 1 shows the word-gram scores of previous challenges. Linear regression gave a trend line with Score = 141.WordCount - 96. As an initial thresholding technique, we examined all results with a score at least 141.WordCount - 250.

2.2 Higher n-gram models for letters

In the development of the software package "AZdecrypt" Oranchak (2023) two users (Jarlve i.e. Jarl Van Eycke and Beijing House i.e. Louie Helm) had built n-gram statistics for n values up to 10. The statistics were stored using log frequency values and scaled to one byte per n-gram. Thus, the 7-gram table used $26^7 = 8,031,810,176$ bytes or approximately 8 GB of memory.

Tables with $n > 7$ store values for only a subset of the n-grams. Support for 7 and 8-grams was added to the AZdecrypt FreeBASIC source from April 2017 (version 1.03). The 7-gram statistics were published in version 1.14 (March 2019) and 8-gram statistics in version 1.15 (May 2019).

The 8-gram models released in 2019 at first stored $125,000^2$ (Jarlve) or $131,071^2$ (Beijing House) values. The most common 125,000 or 131,071 4-grams index the first and last half of the 8-grams. Later, Beijing House released further refined 8-gram models in 2020 and 2024 that each had improved accuracy while also shrinking index sizes to 120,000 and then 106,290 respectively for an ultimate in-memory size of 10.5 GB.

Similarly, the newly created 9-gram models are also derived from the same 10 TB corpus of English text used to create the newest 8-gram models. Their internal structure is slightly more complex and relies on using independent, variable-sized 4-gram indexes (between 46,952 and 106,290) for the first 4 and last 4 letters of each 9-gram, with 26 unique tables for each possible middle value. However, the RAM requirement for these is 88.8 GB and is only intended for "prosumer" computers with at least 96 GB of memory. These values were published in version 1.24 (January 2024).

Finally, the 10-gram values were developed using the most common 300,000 5-grams. The compressed file size is 14,661,017,462 bytes and uncompressed is 90,000,600,001 bytes. These were privately available from 8 May 2024.

We also calculated 8-gram and 9-gram files in a "positional" form specifically for the purposes of this challenge. That is, the first m letters of a sentence starting at a specified offset, and the last m letters of a sentence starting at a specified offset counted from the end. These were calculated for $1 \leq m \leq 8$, that is 16 different positional m-grams.

In addition to the n-gram letter scoring, we also improved the initial 2-word-gram and 3-word-gram scoring using groups of 2 and 3 words. The new 2-word-grams used the most frequent 65,535 English words and different 3-word-grams models used the most frequent 2,000 or 3,400 words in English.

Then after several sweeps using these positional 8-gram models and then adding in scoring from the modest 2-word and 3-word models, although the highest-scoring potential solutions seemed to qualitatively be turning into more natural or plau-

sible potential solutions, there still wasn't a single, clearly distinguishable solution at the top of our scoring lists.

3 Results

3.1 Top Scoring Candidates

On 20 April 2023 the top scoring initial solutions using 8-gram statistics were these.

```
host emotional after Athens
gesture perhaps we can't do it
offense strategy in the sun
you tended the larger units
never described him it shows
hope might have asked aloud
woman said the Pacific poem
```

For the 22 letter solution, the top-scoring solutions were:

```
Marco publishing heresy
firm developed the brief
lured by some of their food
Guy take some of the dairy
unaware line of the major
```

We noted that the "firm developed the brief" solution was already seen online at `https://www.architectsjournal.co.uk/buildings/sarah-wigglesworth-extends-camden-primary-school-through-creative-collaboration`.

3.2 Extended Solutions

On 17 May 2023, we suggested more:

```
Yes had completion in the MiG
West in terms of the said bank
It's okay in the building shop
I cut something up and admire
Belong to this works for some
HIV positive chances Andrew
I stone people who are not big
Case asked for the girls a boy
No it's list of the best change
On orders of the same to vinyl
Spy out of action winter pays
```

On 6 February 2024 we suggested still another 24-letter solution:

```
we are creating a bold pilot
```

The phrase "creating a bold pilot" was seen in several places on the web.

For 22 letters, we found all the following on the live web or archives of the web.

```
would everyone care for it
during eyebrow services
mum marching the babies
the verdict would be half
tell me which publicity
```

More generic solutions not on the web were also found:

```
I do know that I must point
The weapon may be your God
```

Eventually, we resorted to large language models in English to assist with the scoring of complete sentences. This resulted in close alignment with solutions already proposed by Hamidullin on the Schmeh blog for the 24 letter challenge.

For instance, the Hamidullin solution "And perhaps he's collecting" scored very highly.

3.3 Language Models Employed

The "ultimate" English model available publicly on the net was from Buck, Heafield, and Van Ooyen (2014) based on CommonCrawl data.

The authors explained: "We use the 2012, early 2013, and "winter" 2013 crawls, consisting of 3.8 billion, 2 billion, and 2.3 billion pages, respectively." This was 23.62 TiB of English input with 500,262,522,509 unique n-grams (between 1-grams, 2-grams, 3-grams, 4-grams, and 5-grams)

For the 24 letter challenge, "gifts bestowed to her family", "private sector in making his", and "post from other sites using" were seen on the web. However, none of these could be considered complete sentences.

On 11 June 2024 we submitted the email shown in Figure 1 after scoring all possible solutions with two models. We provided files containing the highest scoring million solutions in each case.

The first was trained on "books3" data from approximately 200,000 English books (see Gao et al. (2020)) restricted to a vocabulary of about 140,000 words from "wlist_match7" from Vertanen (2018).

All instances of the letter "J" replaced by "I" in the vocabulary, and only "a", "I", and "o" were retained as single letter words. The "books3" data were then processed with the I/J replacement, resulting in 100.8 GB of data.

Then, a KenLM model "trie file" was built using 5-word-grams.

The trie file was about 163 GB, with 16,031,566,855 different n-grams.

The other was the Buck, Heafield, and Van Ooyen (2014) "CommonCrawl" model pre-

viously described, with a trie file of size 6.06 TB. (Note that the training data for this model did not have the "J" letters replaced by "I" as traditional Playfair cipher plaintext does.)

Despite its colossal size, this model can be run inside a very small memory envelope, and only needs a large SSD to store the uncompressed trie file at runtime. It contains not just 5-grams, but everything from 1-grams to 5-grams, completely unpruned and with Kneser-Ney smoothing. Both those features are key to it being so especially good at precisely scoring and discriminating between so many short algorithmically-generated English phrases.

3.4 Scoring and Filtering Criteria

While the program gave 22 or 24 letter blocks of output, we used word splitting to split them into all possible output sentences (using the 140,000 word vocabulary above). The solver program itself was run in parallel attempting to generate solutions between 3 and 9 words for word-scoring, and then further splitting was carried out on the initial output.

In total, after splitting, we scored about 4 billion possible 24 letter solutions and 1.8 billion possible 22 letter solutions.

The vast majority of these were grammatically incorrect and could not be considered standalone English sentences. Nor did they have an "imperative" or quasi-military tone as did previous challenge solutions.

Schmeh replied that the CommonCrawl model scoring had placed the intended solution for the 22-letter challenge in the top 5,000, and for the 24-letter challenge in top 2,000. He also stated that for the "books3" model results, the 22-letter challenge was in the top 10,000 and the 24-letter top million did not contain the correct solution. (This was due to an error in our splitting code as it chose "BE LOW" rather than "BELOW"). Further, he stated that "both messages might be relevant for a spy".

After a lengthy manual review of the top solutions, we determined that the 24 letter challenge solution was "FIND DEAD DROPS BELOW BRIDGE" and the 22 letter solution was "MONEY IS HIDDEN BEHIND HUT".

"FIND DEAD DROPS BELOW BRIDGE" was found exactly once in an output file produced using (non-positional) 9-gram scoring. The score was 503 and we had used a cut-off value of 500.

> For the 22 letter challenge, the top two in order in both is WOULD EVERYONE CARE FOR IT and PILLARS OF THE INSANITY, and the top 10 of both include AVOID VISUAL DEFICIENCY (a phrase often seen on web). In the top 20 both have THE VERDICT WOULD BE HALF (seen on web).
>
> For the 24 letter challenge, the top 20 of each both have AFTER THAT ABOUT CLOSING UP, GIFTS BESTOWED TO HER FAMILY, IN EVOLVING THE WORLD MUSIC, ITS HUGE PERFORMANCE MAGIC, THEY SAY THE LIQUOR HAD GONE, and WE ARE CREATING A BOLD PILOT.
>
> The top solutions for 22 and 24 in the books3 model WOULD EVERYONE CARE FOR IT and AND PERHAPS HES COLLECTING by a large margin. (APHC is line 84 in trie 24 solutions)
>
> A bit further down 22 solutions I like WAIT COME BACK TO US FIRST (line 848 in trie 22, line 102 in books3 22). It has a kind of imperative nature the other larger challenge solutions had, plus very simple words.

Figure 1: Email submitted 11 June 2024.

In contrast, "MONEY IS HIDDEN BEHIND HUT" was found 34 times in output files.

We used the CommonCrawl and books3 model, and the various n-gram letter models to assess the relative difficulty of the 22 and 24 letter challenges compared to the 26 letter challenge.

We determined that the 26 letter Playfair challenge solution "WAIT FOR FURTHER INSTRUCTIONS" was the top scoring solution in each KenLM model and occurred in the top 25 solutions of every n-gram letter model with $n \geq 4$. It was the top solution in every n-gram letter model with $n \geq 6$. In contrast, the 22 and 24 letter solutions (considering the highest-scoring million solutions) were never in the top 50,000 highest-scoring solutions. Thus hill-climbing or simulated annealing approaches designed to maximize n-gram letter scoring did not assist much.

3.5 Alternative language models

We had also tested the scoring from another large model we found online (described in Jansen et al. (2022)). This model available as a 10.8 GB binary file. It was trained on harmful and toxic language. The potential solutions were ranked in a quite different way to the other models. For instance, for the 22 letter solution, "I am here to convey the porn" and "I am seeing a lovely cunt" were high

scoring, and for 24 letter solution, "pin up with the naked booty" scored well. Thus, if a model could somehow have been trained on "spycraft" or "military" text only, with an imperative tone, it may have done well with the challenge.

3.6 Assessment

In retrospect, it seems that proposing a challenge with missing articles and connector words made it nearly impossible to distinguish between very common potential solutions and the intended correct solution for such a short cipher. By this we mean that the intended solution with connecting words (i.e. "Find the dead drops below the bridge" and "The money is hidden behind the hut") would have scored very highly indeed, but not met the length constraints.

4 Conclusion

The 22 and 24 letter Playfair challenges proved to be much higher in difficulty level than the previous challenges which ranged in length from 26 to 50 letters.

Using previous techniques such as n-gram letter and word scoring, even with values of n up to 10 and incorporating all additional techniques (tabu search, multithreading, simulated annealing) and new techniques (positional n-grams) was not sufficient to bring the intended solutions anywhere near the highest scoring results.

For the 24 letter challenge, the intended solution was seen only once in more than a year of trying different scoring methods.

The Parker Hitt quote about cryptanalytic success requiring "perseverance, careful methods of analysis, intuition, and luck" (in order) proved true here. This was especially true for the 24 letter solution.

In practical terms, it was not possible to identify the intended unique solution, even though for the 24 letter solution, the length exceeded Deavours' estimated unicity distance for the Playfair cipher in English.

Another practical consideration for challenges of this nature in future is that the models used to assist solution may well now have been trained on AI-generated content, which may render them less suitable for human generated challenges. Projects such as "Paracrawl" of Bañón et al. (2020) attempt to improve upon previous crawls while avoiding these pitfalls. The cost of training a new model with such data would now be substantially lower than for the 2014 CommonCrawl model used here, assuming sufficient compute power was available.

5 Acknowledgements

We would like to thank Klaus Schmeh for constructing the challenges, and the University of Queensland High Performance Computing Resources for use of their "Bunya" supercomputer.

References

Bañón, Marta, Pinzhen Chen, Barry Haddow, Kenneth Heafield, Hieu Hoang, Miquel Esplà-Gomis, Mikel Forcada, et al. 2020. "ParaCrawl: Web-scale acquisition of parallel corpora." Association for Computational Linguistics (ACL).

Bean, Richard. 2020. "The Use of Project Gutenberg and Hexagram Statistics to Help Solve Famous Unsolved Ciphers." In *HistoCrypt*, 31–35.

Buck, Christian, Kenneth Heafield, and Bas Van Ooyen. 2014. "N-gram counts and language models from the common crawl." In *Proceedings of the Language Resources and Evaluation Conference 2014*, 3579–3584.

Cowan, Michael J. 2008. "Breaking short Playfair ciphers with the simulated annealing algorithm." *Cryptologia* 32 (1): 71–83.

Cowan, Michael J. 2009. `http://www.mountainvistasoft.com/cryptoden/index.php/algorithms/churn-algorithm/18-simulated-annealing.html`.

Deavours, Cipher A. 1977. "Unicity points in cryptanalysis." *Cryptologia* 1 (1): 46–68.

Dunin, Elonka, Magnus Ekhall, Konstantin Hamidullin, Nils Kopal, George Lasry, and Klaus Schmeh. 2022. "How we set new world records in breaking Playfair ciphertexts." *Cryptologia* 46 (4): 302–322.

Gao, Leo, Stella Biderman, Sid Black, Laurence Golding, Travis Hoppe, Charles Foster, Jason Phang, et al. 2020. "The pile: An 800gb dataset of diverse text for language modeling." *arXiv preprint arXiv:2101.00027* .

Jansen, Tim, Yangling Tong, Victoria Zevallos, and Pedro Ortiz Suarez. 2022. "Perplexed by quality: A perplexity-based method for adult and harmful content detection in multilingual heterogeneous web data." *arXiv preprint arXiv:2212.10440* `https://huggingface.co/oscar-corpus/harmful-kenlms/blob/main/en.binary`.

Kopal, Nils. 2018. "Solving Classical Ciphers with CrypTool 2." In *HistoCrypt*, 149–010.

Lasry, George. 2018. *A methodology for the cryptanalysis of classical ciphers with search metaheuristics*. Kassel university press GmbH.

Lasry, George. 2019. "Solving a 40-letter Playfair challenge with CrypTool 2." In *Proceedings of the 2nd International Conference on Historical Cryptology, HistoCrypt*, 23–26.

Lyons, James. 2012. "Cryptanalysis of the Playfair cipher." `http://practicalcryptography.com/cryptanalysis/stochastic-searching/cryptanalysis-playfair/`.

Oranchak, David. 2023. "AZDecrypt." `https://github.com/doranchak/azdecrypt`. [Software].

Schmeh, Klaus. 2019. "Klausi's Krypto Kolumne." `https://scienceblogs.de/klausis-krypto-kolumne/`.

Vertanen, Keith. 2018. "Wordlists." `https://www.keithv.com/software/wlist/`. [Wordlists].