Unit 178
Attacking Enigma with the index of coincidence

Analysts during World War II used every bit of their computing power to break the Enigma. In the modern world, we have so much computing power that we waste most of it. A statistical attack like the one we present here uses a lot of computation. We are going to use the index of coincidence (IoC) (see Unit 12) to help us find the configuration of an Enigma machine that was used to encrypt a ciphertext. In this unit we will restrict ourselves to a three-rotor machine, but the method can be generalized to four-rotor versions.

The attack has these three stages, which we will discuss in more depth below.

1. Try all choices of reflector and rotors and (provisional) initial positions and keep the configuration and starting positions with the highest IoC.
2. Keep the reflector and rotors from stage 1, but vary the ring settings and initial positions and keep those with the highest IoC.
3. Try plugboard wirings one pair of letters at a time and maximize the IoC.

In the first stage, we try all possible combinations of reflector and rotors. Remember that a rotor cannot be used twice in the same configuration, since each machine was outfitted with one of each. We also try all possible starting positions for the rotors (the message key). In this stage, we fix the ring settings to be at 1=A for all rotors (we could take any arbitrary ring setting, but it should not matter much). There are three reflectors, eight rotors, and twenty-six possible starting positions for each rotor, so we have to try $3 \cdot 8 \cdot 7 \cdot 6 \cdot 26^3 = 17{,}716{,}608$ configurations, and therefore we recommend a compiled programming language such as C, rather than a slow interpreted one like Python. The configuration whose resulting plaintext has the highest IoC is kept for stage 2. The winning margin may not be large, so we should take some care; if there is no clear winner, perhaps the method has failed and we need to try a more exhaustive search (see below).

We consider the initial positions of the rotors found in stage 1 to be provisional because the true ring setting can affect whether the rotor acts more like it starts in one position or more like in a neighboring position. So in this stage we will give some leeway to the rotor positions, and allow it to have a starting position one place earlier in the alphabet. The ring setting on the leftmost rotor is irrelevant for a ciphertext-only attack because it only affects the locations of the stepping notches, which aren't used for that rotor. This ring setting is important for cipher clerks, because it affects the message key. But for us, we can fix the ring position to 1 and not allow it to vary. As we increase the ring setting of a rotor, the starting position increases with it. If, for example, in stage 1 we found the best starting position for the middle rotor to be G, then we should try these possibilities for that rotor (remember we give it some leeway):

ring setting 1	starting position F
ring setting 1	starting position G

ring setting 2	starting position G
ring setting 2	starting position H
ring setting 3	starting position H
ring setting 3	starting position I
⋮	⋮

We need to also give leeway to the starting position of the leftmost rotor, even though we do not vary its ring setting. So there are $2 \cdot (2 \cdot 26)^2 = 5{,}405$ possibilities to try. The one that gives a plaintext with the highest IoC is kept. In this stage, there should be a clear winner.

The action of the plugboard causes more apparent randomness for the cryptanalyst. But for each pair of letters swapped by the plugboard that is correctly identified, the randomness can be reduced. Hence, every time we account for a pair of letters cabled on the plugboard, the IoC of our plaintext will increase. To find the first one, we try $26 \cdot 25 / 2 = 325$ pairs (divide by 2 because AB is the same as BA, etc.). We keep the pair that gives the greatest increase in IoC and include it in all future decryptions. To find the second pair, we try $24 \cdot 23 / 2 = 276$ pairs (we don't include the letters of the first pair in the search). When we reach a point where adding another pair to the plugboard reduces, rather than increases, the IoC of the plaintext, then we stop. At this point, the attack is finished.

We must, of course, look at an example. Consider this ciphertext that was encrypted with a three-rotor enigma. It is long enough that we are confident of success.

```
NNBOFZRDRLRSYRRVJPSXPSZMWHUHESXPFKKHYRNUFYMYXQJPMOLTPMRTKKCVJQNT
HAPVQBFHJWVRXWDOMQNWRJOJXIPBBHZVLPSGZJCTVUMMUUEFHEKNYHWIXVCJQKSJ
IEZLSUXZLRLKJBOYCZHNYUCKWPAIYCMUTKUFHMCOXWLEMSYWKZYBOFZCPTQZKNZQ
DDGTPHRMFREBNMQUFMRELRGAFXVRWKAOEVKBQIMGMTIBKBGLRRAYBOIRXEQLFOZA
QTDFDNUAGDHVATDTBSOXAFTHZAHWKAWACQKLRKCJTNQEWODQJEUMNQXKCAYQVVBL
KMDQNIMEHKVMHENUTIPTAPNCQFPHVKEJLDEIDGXGRYJAZJILHLJUEJNYBFTUPEMP
FXUXBICDFBUDYLMJMDKTUHWKKLOKJNAXRIUDIHVSDJTSTRRNNZKIAVEYASORARRS
TKUWJNTFYCSOSQEFGZKJPFDDHPXSUJTRFSCZPJKPZVQLNNTGODWCTCKEUBJWXNNA
WFBJUQXEDYIVTYJMNLSZQQMQVKVNNORUFPZPXTHZQIDHTFDYYXVJMAXLWDWHMUWS
BKTFYMJPKGNGILAUPPVXNWLUDAISQPLZUZDGXRFOHUYJRRBHWIFHVBJQXBGVAFTT
IZGXXZVGOEOUUUXSUXJKMBSEPTLOBZITMNIKYQBHMVIIHAOJNILNNIKCEKERRAOY
YNPWLCYQRGUALYOZFUZVSHXDGKLSDRLFZVKPDZVMKUHMJNMZXBSWLZGIURVJIQZX
STUVXIXLDHOKNEDJLXCKMEMDWVWAMPZZZOBZYVHMDNCJHXCOJHZUHEPYWHKRJMVH
STVUPGRZEJZBTOJRCTQFYLOQFCWMDDBKHRKUXXUOVYBOQZTGJBQKDBQLENNNXBOV
OTSOVMSXGIMIGXFIPMEFFUAWUYZBQBEAXJFDWBDUAAPRYRXNYBMIH
```

In the first stage of attack, we try all possible reflectors, rotors, and starting positions. Ring settings are all set to 1. We sort from lowest to highest IoC and find that the best is when the reflector is B, the rotors are II, III, V, and the starting positions are KEZ. It's a narrow victory, but we will run with it. If we find that we cannot solve the ciphertext, then we can always come back and try the second-place configuration.

reflector	rotors	starting positions	IoC
⋮	⋮	⋮	⋮
A	VI IV VIII	IZS	1.0603

C	VIII II V	QHT	1.0607
B	V VII VI	NIK	1.0625
B	II III V	KEZ	1.0635

We keep the choices of reflector and rotors, and now concentrate on the ring settings. For the fast rotor (on the right), we found our result above for ring setting 1 and initial position Z. So we try 2 and A, 3 and B, ..., 26 and Y. But because of the uncertainty we have in the starting positions, we try these plus or minus one step. At the same time, we try varying the ring setting and starting position in a similar fashion for the middle rotor. For the slow rotor, to be safe, we allow the starting position to vary by one step, but still keep ring setting 1. After the search, here is the end of our list, sorted by IoC:

ring settings	starting positions	IoC
⋮	⋮	⋮
1 17 7	J U F	1.2585
1 16 8	J T G	1.2659
1 17 9	J U H	1.2833
1 17 8	J U G	1.2943

The best is for ring settings 1, 17, 8 and starting positions JUG. We continue with the assumption that we have found the correct settings for the rotors and now turn (pun intended) our attention to the plugboard. There are (26 · 25) / 2 = 325 ways to connect two letters. If we decrypt the ciphertext with our favorite rotor settings and each possible choice for one wire on the plugboard, we find these IoCs:

plugboard pair	IoC
R V	1.1833
K R	1.1899
⋮	⋮
C F	1.3799
P T	1.3888
B X	1.3912

The last few are all good choices, but to follow our method, we keep only the best pair, BX, and repeat the procedure assuming that this pair is correct and look for a second pair. This time, we only need to consider (24 · 23) / 2 = 276 new pairs.

plugboard pair	IoC
R V	1.2652
I V	1.2686
⋮	⋮
C F	1.4838
A L	1.4876
P T	1.5027

So we keep the pair PT and continue to find the next pair to be AL, giving an IoC of 1.6081. We are now in the range of typical English text, so let's pause for a moment and look at the resulting plaintext:

```
ZEENTRMAGEGINMAOCMOGSIEURBOURDOGRSMAIGHYXGOBJEFTIVEIMPARTIHLAGSI
MPERSOAALSTUDYAGD...
```

OK, we're not finished yet. We can see some words peeking out at us, but it is clearly not correct. So we find a fourth pair, CF, with IoC 1.7200. Let's pause again and see the plaintext:

```
ZEENERMAGEGINMAOFMOGSIEURBOURDOGISMAIGHYAGOBJECTIVEIMPARTIALAGSI
MPERSOGALSTUDYAGD...
```

There is marginal improvement. Searching for a fifth pair reveals GN, with IoC 1.7960. Trying to find a sixth pair only *reduces* the IoC for this ciphertext, so we may be at the end. When we check, the plaintext now looks correct:

```
THEGERMANENIGMAOFMONSIEURBOURDONISMAINLYANOBJECTIVEIMPARTIALAND
IMPERSONALSTUDYAND...
```

We have found the entire key. The plaintext is from *German Problems and Personalities* by Charles Sarolea:

> The "German Enigma" of Monsieur Bourdon is mainly an objective, impartial, and impersonal study, and the author has been careful not to obtrude his own private views. It is only in the last chapter that he attempts to draw the lesson and point out the conclusion of his own inquiry. And his conclusion is an eloquent though restrained plea for a Franco-German *rapprochement*, and in favour of the only policy which will bring about that reconciliation. France, he argues, does not want a revision of the Treaty of Frankfurt. She does not want compensation or revenge. ... A French statesman, on the eve of the Treaty of Frankfurt, made the rhetorical statement that France would never surrender one stone of her fortresses nor one inch of her territory. Animated by a very different spirit, modern French statesmen do not claim back to-day one inch of lost territory. All that the French people demand is that the claims of justice shall be heard, that Alsace-Lorraine shall cease to groan under the heel of an arbitrary despot, that Alsace-Lorraine shall be governed according to her own laws, that the Alsatians shall be treated as a free people, and not as conquered subjects.

In the case where too many cables are used in the plugboard, it may be impossible to succeed in stages 1 and 2. When that happens, we need to try all configurations of reflector, rotors, ring settings, and initial positions of the rotors. For each, we try to maximize the IoC in the same way as in stage 3. Only when we find an acceptable plaintext do we stop the search. This can be a long and difficult process, so it would be best to use a fast programming language, such as C.

The index of coincidence is not the only statistic that we can maximize to break Enigma ciphertexts. See references for refinements of the attack and the use of other statistics.

Reading and references

James J. Gillogly, “Ciphertext-Only Cryptanalysis of Enigma,” *Cryptologia* 19:4 (1995) 405-413, https://doi.org/10.1080/0161-119591884060

Note that there are 325 possible two-letter plugboard wirings, not 625.

Heidi Williams, “Applying Statistical Language Recognition Techniques in the Ciphertext-Only Cryptanalysis of Enigma,” *Cryptologia* 24:1 (2000) 4-17, https://doi.org/10.1080/0161-110091888745

Olaf Ostwald and Frode Weierud, “Modern breaking of Enigma ciphertexts,” *Cryptologia* 41:5 (2017) 395-421, https://doi.org/10.1080/01611194.2016.1238423

Programming tasks

1. Implement the attack on a three-rotor Enigma in which it is possible to do stages 1 and 2.

2. Implement the search on an Enigma that has too many plugboard connections for you to succeed with stages 1 and 2.

Exercises

1. We can start off with something easy. This ciphertext was encrypted with a German army Enigma without a plugboard. The army used three rotors chosen from only I, II, III, IV, and V, and reflectors B and C. To increase the IoC, X has been placed between words and XX between sentences.

```
XBCWHIYYBMIUFWJBFMBNCYRABCJABKUHIWUTWUYTIDYMWYHZQWOVFGUQELLAZDGY
BGNYEIWXTHISOPYETEASUBYSVJRDBPEAIWSBBPCQGRRHZDNFMQCOCPISCFNIXZCJ
LCEBZFKDDBOQRHWAEGOFOKHAXVCSVWTPQHYBVTAEIMXHWMUZZLUYRKNCTJEEZXBI
KAUCSBGLDDTPVLFWQXTZRTYBVOYJVXXGWJDZXOEHYPTKTOIPSEGCLWFMOKSABZWD
ESGNSPEFGKNYFIPPJFWTRTKYOSLLGPHLWRJUBVLEFDRDCUIWGYURLWNUPTSHDPOV
MIYDWFXLOPDFAZRPQGWEBEILBDNDCDKDRLHTCIYNOMRYLVZZFNWDHVTBWDXMRHUE
SEEBPOYLJVZXVCACJYZJYSHYAUPMOOHBFYBDSTSOXUVETTSXLSSPLRHRTARRIAVO
QKCRKGWEGXQAUAVSZEDPIAJUNSARRQNJZUWQHOBSTYAILRLECLWUABQJHPROOCPE
ZUJYWIJZMVQGGCEBVHOVGHVVPCBNVVNNXXASQHMKPSZPWLRMMW
```

2. Break this ciphertext that was encrypted with a three-rotor Enigma. The plaintext is in English. To increase the IoC, X has been placed between words and XX between sentences. There are five cables in use on the plugboard. The rotors were chosen from all except beta (β) and gamma (γ). The reflector is B or C. There may come a point where you have to try two forks in the road.

```
CCFZIAPYHFQWOWJULWBHWFGIXPYDRGEDDSBXCGVXLQCPWMXEMYRWJKFUQPFGDESI
PRJIIWBISATECJCFRNAXWCDKQPDDRSPMHACXVJYZLZKABALXTSWISWVCLJNIZOAK
LVLNGHCHFPMLBGGCPKROXBWDUFCRHBNLNJHJJEGVFQNSAOQUZESPDGMHJVCRMAUX
KQRMWQQFBOKDNOANLHTGGHEGVDJGERNNQMFBVPNSWTYPRLPADUCYPNGWBCCQRDRK
MFAZTXICVIAGNBXKZTCWATGREROIDTDENCZVEUTJIIGGXBEVPBAUEMEKPUWNRLNF
MIJXCAYKVKCKEHKMFKJADUSMCYBTVNLHWFKNPXEHQOGUHMJDRIOBQEWRSMCGSVCO
VYPJLMJCNCEVXBAPKTERFWAKDPVVCZBGXYLUHKTEIJCYRQZZLCPCNFLFHGMSDAUK
```

```
UIWJIVNBRHKRUMNGTBUMOGKJFZTNZMOUKFKCJVKAXNRPVRLDQUMLVMIMFVKGTPRK
TPPZXQUHXKOFDQFZPRBHWOFBUYJDPDQDHLGZKEHRSVRTEGJBWPMRTTZYZHAUAGTR
EOUPMTGLHOHMAKIAUQPUOOIWIGBMGECRQZQVWVBBCLGDDNSAMYMUAAERFVNPVINT
YGSKTYXTCPTBURVRGIXEKIQQMWLWYDOMAAKBUNPEZYMPLMPZVAKKAAFBGGMYQRRA
PDUNFPROCZBBUDJVHGYYHXOAWOZCVLFUVTELHKFGWBKZPTCTXJFLUKVDHXVKOWTK
BNEUAEYMHLGEPILRCVPWEGLIOBIUUZWMEJPPIDJTCRXJVBQUZRHQTBYDEVEFCOLH
JWVGBBSJDEDLGKQSIVVLSTZWBNHLKCNUKSUSFEWIHEDGYTFWWJKHNZKCEMYOZGKV
QUIPGGPSDZALLMDMJUYWIGUSFQUQVPQYZCTMUQTYBAZGTUJFFWQTHJGHUIAGRGSN
ZNBYNLNTZPJJGMDFTWBHUQTYFGVMVPYBNTAODQFSQPYLLDZGYHAGZJIWKTDEORYP
AIHTVDOIZYXSTUWKBLIGCCQTZJUQVEPIEUTTVPLKZJZNYOSTUXYEJJFKZFLOUAWN
NWGXCMYCCRDAGONYQBYBGOJMQPPGIYTRFCVXSZVOTFKNGGCPVLVSPQRBFUJFZKPE
ZHYKTPVWVFBUBIFLQJMLTHBTKE
```

3. Decrypt this ciphertext that was encrypted with a three-rotor enigma. There are ten cables in use on the plugboard. The reflector is C and the rotors are II, I, V. The plaintext is in English; believe it or not, the word “enigma” does not appear in it. Remember that the first ring setting is irrelevant.

```
UOPIIKUDEZGWQFXUVYSTROHVNAYLAKHPJVRIVEPNPTTMHXGJYFWJQPAWMZVQFPYB
FZYMYJKOFROAABKIXSFBOYFWTHVNYMWEJKYWVCHCJMFHORYLNOLLIUYNZETGGZOQ
UUANZXKQWLNYNZEXDHWAKXPXLMDULPTANBGAYTWOBLYVGVFSNQDEQQGVZAMEGADB
JPNYRFVUPEJTZKMKUUJUHOHFJNTBLACTDJFRQTQFPTGWSXMVSRIAKVOJGOQTRZGX
ACYYRZQMSRBOMNRDZGDKYQZJMLOOWGIIBIUMHYBMBXSFSLYQGBWXYUDAPWDGYARW
JGMTDXNDYBYRNZFWTAPLHBZPZADJDBZODAYGTKQCZIZQLKYUFXRDXZJEVYEUCZCF
MUGYMTRCQXNMJGHWYDYLERSXUPNCZICUZHSAPPBTZMTDUWNOKEKRRESQJIKKLQVP
VAUJHXTQJAIOVBOSPENPMWHFERFAWJDLISNIFFGTBBFYCBFJSXOFOZYMWDMPEBIY
VRTDTHKYSXDYGRALISAJBWFQXKVYWTYFMORTYDVPKYXFXWYEOSEWGPIZOQNBPYQH
DVNDNQYRJCAVXZKZIXZEXDSGBEVAVQGTGPSJUUXNTISCYPRMNZWCXVWLRUTBEXYG
UNLBKASRUPIFXKWXCPWQTVIMRHBCMNRSOVGWZZQBFPEVRWCVRGSRMAJBQXLVGEGL
HQRULPWLXYOHPKZMLIKPCZASULNRWJRFMYVVTFBXBPICXHZCOGZQEXPSSOLUIIXD
ZM
```

4. Decrypt this ciphertext that was encrypted with a three-rotor enigma. There are ten cables in use on the plugboard. The plaintext is in English. This could take a long time.

```
HQWSEQGSTBGBHDBGVNYJMZHFZNIFOCTADLAHXYJXCHMQHWMEAXVZGHBMLDJGELOG
DUVVJDHWUHIYAEVCWHQSCLYHVOTKYXMOLYRIDAVVJXYGEBJUYCBKREATBYVWRBQI
EJRARBSXELJJFVJOMSJUDTKIKGQQTUINRAZLGXRTEFSDGDMNFYVAGARQNXLMAQVH
TFNVHCGJXVQQIRGKCGUWHVFZXSQRHKNJCLTKDXZZOJFWXUNLJEAPDSUXNOLDWNNR
MQMOZOOVLKRMGZVTMMKNHZIZXUIQKJPDRQSCBUFRKKDPPPIRPZBDFUOLNOPBOMGV
RFIZXWONCNFKYXPGBDOPKBQDIPLBFSDCLVMAVDFGQPMKYPCSDVKUQBHZPQRHBUZG
LMNTXVGZFFQWPOUFMCUCHLDMGLIHYIBKBDRHBNZDVZYTJJRIGYMDBEHGKCUZPTQW
IDJTZVMWXODOKUWBBGKRWPXHEWJOLEWECGCPDKATIYTJCENQKKTOLCJDIPIZCZPA
CEVIDGSSXLIQMGHLOLT
```

If you enjoy breaking Enigma messages, there are some challenges (with ciphertexts in German) at this site: https://www.ciphermachinesandcryptology.com/en/challenge.htm But note that in Enigma messages, the Germans put `X` between sentences, `J` around proper names, used `Q` for `CH`, tripled letters in designations such as U-boat `U235` becoming `UUUZWEIDREIFUNF`, and, worst of all, wrote in German.