Dawn of Mathematical Cryptology: Probabilists vs Algebraists or Algebraists & Probabilists?

Marek Grajek
freelance consultant and historian, Poland
mjg@interia.eu

1 Introduction

Traditional cryptology, before the advent of the ciphering machines, relied mostly on the linguistical methods, and the role of mathematics in the codebreaking was limited to counting the frequency of letters, their pairs and triplets. Machine cryptology changed everything; only mathematicians were able to interpret even the bare numbers of combinations resulting from the use of the ciphering machine. The first successful application of advanced mathematics in cryptology, Marian Rejewski's success with Enigma, marked a change of paradigm; his attack was based on the algebra and the group theory. However, soon after the outbreak of WW2 the Germans had changed the Enigma operational procedures, rendering most Polish methods of attack ineffective. British mathematicians, who took over from the Poles, had to revert to the old and proven methods based on probability and statistics, which dominated their work during, and well after WW2. It was only 30 years after the end of this conflict that the role of the algebraic methods was restored.

This paper presents the early period of the development of the mathematical cryptology, focusing on the clash of two approaches to the codebreaking; that based on statistics and probability on the one side, and algebraic methods on the other.

2 Historical context

Although traditional, historical codebreaking has always been based on linguistics rather than mathematics, at least since eight century a component of simple application of math was present therein. Al-Kindi, an Arab polymath living in Baghdad in ninth century, described in his "Manuscript on Deciphering Cryptographic Messages" an attack on the monoalphabetic substitution ciphers based on the natural frequency of letters in the language of the clear text. As far as we know his work was based on the earlier (and presently lost) writings of Al-Khalil[1] (also known as Ahmed al-Farahidi), living in Basra in eight century. Their works linked early cryptography with the equally early methods of mathematical statistics. It should be noted that Al-Kindi's interest in statistics was not limited to the secret writings. He proposed also a statistical approach to the medical treatment evaluation.

During the mediaeval and early modern periods attacks based on the letter frequency were still popular due to the popularity of the nomenclators; monoalphabetic substitution representing a part of the nomenclator made it vulnerable to statistical attack. Later on, when the codes and nomenclators started to lose their popularity in favour of simpler and more practical ciphers, statistical attacks have gained in importance; codebreakers started analysing not only the frequency of the single letters, but also their pairs, triplets and entire, popular words.

Invention and fast adoption of the telegraph has changed this picture for a moment. Early telegraph required not only hiding the message content, but also its compacting. Codes provided an easy and practical answer; second half of the nineteenth century was heavily dominated by the use of codes, which, from the codebreaker's perspective, required the application of the linguistical rather than statistical methods of attack. Use of the radio during the Great War has brought another game changer. Both sides used radio on a mass scale. Ease of interception of the radio messages forced the application of cryptography at the equally mass scale, and the traditional codes were getting more and more impractical; during World War One ciphers replaced the codes as the mainstream of cryptography, and, consequently, statistics replaced the linguistics as the mainstream of cryptanalysis.

However, statistical methods used in cryptanalysis represented rather elementary applications of mathematics, which could be dealt with by amateurs. Immediately after the end of World War One agencies of major countries dealing with cryptology did not realize the need to employ or train mathematicians. If some mathematicians happened to be employed in the crypto world, it was only due to their general intellectual discipline, and not to the particular skills resulting from their scientific discipline. Werner Kunze was employed at the German foreign ministry cipher office in 1919, but it was only in 1936 that he became the head of its newly created mathematical section. William Friedman published in 1930 an offer to employ three “government mathematicians” at some obscure agency of the US Army. From the memories of Solomon Kullback, Frank Rowlett and Abraham Sinkov, whom he selected from among the candidates, we learn that for the first several years nature of their occupations was rather distant from mathematics.

It seems that the first agency dealing with cryptology that consciously and purposefully decided to employ and train mathematicians was the Polish Cipher Bureau in 1928. Effects of that decision are well known among the historians of cryptology; after the half year training in cryptology in Poznań, in 1929, and three years long period of apprenticeship in the codebreaking, Marian Rejewski was asked to take a look at the real objective of this effort – the Enigma cipher. It took him less than three months to break the cipher and, simultaneously, to change the nature of cryptology forever.

3 Probabilists vs Algebraists

When in October 1932 Maksymilian Ciężki had asked Marian Rejewski to take a look at the materials that Polish Cipher Bureau was able to gather about Enigma (Rejewski, 1967), Rejewski, in spite of his over three years long training in cryptology, was still a mathematician rather than the codebreaker. One might say, luckily for the civilized world; had he been the cryptologist, he would have probably tried to apply the traditional codebreaking methods, completely ineffective versus Enigma cipher. Rejewski started his work identifying some purely mathematical features of the cipher and continued transforming his entire knowledge about the machine and its cipher into a system of equations. He was unable to solve these equations outright, as the variables they contained represented unknown permutations rather than the numbers. Theory permitting to solve such type of equations was missing and Rejewski had to provide it himself, which he did, and in the last days of 1932 he managed to solve his equations, reengineering thus Enigma machine in a purely mathematical way.

Terms used in the description above do not leave a shadow of doubt that Rejewski was using an absolutely pioneering approach. System of equations represents a term functioning in the purely algebraic context. Permutations are used in the context of the theory of groups. Neither reminds ideas or notions used in the probability or statistics, dominating codebreaking up to that moment. It is somewhat surprising that Rejewski had not started his attack from the statistical approach, considering his earlier professional plans. Just after having completed his studies in mathematics at the Poznań University, he decided to continue education in the actuarial statistics at Göttingen. One of his relatives was among the founders of the first life insurance company in Wielkopolska, and Marian Rejewski obviously planned to start a career in the insurance business.

Algebraic and group theoretic approach, used by Rejewski in his breakthrough attack at the Enigma cipher, had numerous advantages over the statistical attacks used against the earlier hand ciphers. Its crucial advantage was that it worked. Codebreaking agencies of the major countries initially declared helplessness when confronted with the Enigma cipher. William Clarke, one of the veterans of British Room 40, remarked in a memorandum written in 1937 that “only one cloud obscures the horizon – possibility of general application of the ciphering machines. One can argue that it would mean the complete end of the codebreaking”. Frank Birch noted an opinion of one of G.C.&C.S. heads of section stating that “(a)ll the German ciphers are unbreakable. (…) putting pundits on them represents a waste of time”.

Rejewski’s approach was unique among the traditional codebreaking methods, as its success was deterministic rather than probabilistic. Most codebreaking methods used up to his breakthrough were offering only a promise of success, without granting it. Success depended on many factors beyond the codebreaker’s control: errors committed by the cipher clerk, external evidence permitting to guess the content of the message, inspired guess of the probable cleartext. Algebraic approach invented by Rejewski virtually granted the success, provided that the codebreaker was able to accumulate some 80-100 messages during a single day.

Finally, with the proper tooling Rejewski's method was extremely efficient and fast. In 1935 Polish team decided to construct a simple electromechanical device named cyclometer. Cyclometer was used to simplify the preparation of the catalogue of so called cyclic characteristics. Ready catalogue of the cyclic characteristics permitted to break over 70% of the intercepted German messages within just few hours after interception. In many cases the deciphered messages reached the eyes of the Polish intelligence officers before they landed on the desk of their rightful German receiver.

Success reached using the algebraic approach did not make Polish mathematicians blind to the possibilities offered by the traditional statistics. In fact, the team seems to have been divided between the adepts of algebra and group theory and those of the statistics. Jerzy Różycki, the youngest member of the team, from the very beginning was focused on the statistics, usually with great success. Still during their apprenticeship the team was asked to break the training code of the German Navy. Różycki started his work with the observation that in any language number of words starting with any particular letter of alphabet represents characteristic feature of the language. He divided the intercepted codewords into the groups of various frequency and started thus successful recovery of the codebook (Rejewski, 1967). A little later Różycki invented the ingenious method permitting determination of right-hand Enigma rotor, called the "clock method" (Rejewski, 1967). His method relied on the idea of the index of coincidence, originally described by William Friedman in 1922. There are some indications that Różycki might have discovered the index of coincidence independently of Friedman's original work (Grajek, 2019).

Algebraic approach served the Polish team well until 1938, then the situation started to get complicated. During the Munich crisis German army has modified Enigma's ciphering procedure, making cyclometer and the catalogue of cyclic characteristics obsolete. In the increasingly confusing political situation the codebreakers had to find a new way to break the cipher, and to find it fast. Rejewski (1967) responded with a concept of the "bomba" – an electromechanical device running through the entire key space within less than two hours and stopping whenever potential solution was found. He developed the new idea within a month and it took the AVA company working for Polish intelligence service another month to deliver six prototypes, but nobody was proud of this achievement. First – because in December the Germans increased the number of rotors to five, increasing tenfold the number of bombas necessary to break the cipher. Second – because Rejewski seemed to consider necessity to reach for the machinery as the failure of his beloved mathematics. And third – because the bomba did not implement the attack based on the algebraic, but only statistical approach.

Most Enigma historians assume that bomba was designed to look for so called females, i.e. one letter long cycles in the Enigma cipher, transforming some letter of the cleartext into the same letter of ciphertext twice in the distance of exactly three characters. The very idea of females was valid only in the context of another method of attack, being developed in parallel to the bombas by Henryk Zygalski, and therefore referred to as Zygalski sheets. The females in the Zygalski sheets represented the cyclic property of the Enigma cipher and their existence and nature resulted directly from the algebraic considerations regarding the cipher.

This was not the case with the pairs of letters sought for by the bomba. In his description of his construction Rejewski (1967) referred to the object of its search using the term "spectacles" rather than females, stressing the difference between both concepts.

Spectacles represented a purely probabilistic property of the cipher and therefore the bomba did function only in the probabilistic and not deterministic fashion; under certain circumstances it could find the key to the cipher, but the solution was not granted. Rejewski never openly demonstrated his disappointment with his own idea. However, an emphatic reader may easily spot the difference in the tone of his description of bomba and purely mathematical methods of attack. Describing the bomba Rejewski pretends to have forgotten the details of its construction and functioning and attempts to diminish its role, revealing involuntarily his emotional attitude towards his own creation. His remarks provide a strong contrast with his comments regarding the Zygalski sheets which, although they do not represent his own idea, belong to the mainstream of his algebraic thinking about the cipher. It is a pity that even writing his memoirs in the late 1960s, he remained ignorant that his bomba represented the foundation for a family of machines constituting the basis of the Allied cryptologic effort during the war.

It was true, however, that the algebraic approach preferred by Rejewski and his colleagues has reached its apex sometime in 1937/1938, and was doomed to decline over the next few years: events they were able to keep under control since 1932 started to slip out of their hands. Everything started from the changes introduced by their adversary around the Munich crisis. Members of the Polish team used to comment mistakes made by the German crypto service saying "they'd better do it in this or that way…". And surprisingly, in just few months their adversary was changing his systems strictly following their own opinions. Poles started to suspect the existence of a mole within their closest circle (which, according to our present state of knowledge, was not true).

One of the conclusions of the Pyry meeting in July 1939 divided the efforts between the cooperating parties; British codebreakers were responsible for the construction of the necessary equipment, French for using their agent in Berlin to get more information about Enigma, and the Poles for the studies in the theory of Enigma ciphers. That division soon fell victim of the wartime reality. Polish team was able to find a refuge in France and to reorganize in P.C. Bruno only to discover, that the Poles represented virtually entire cryptology of the French army. Concentrating their efforts on the daily, operational codebreaking they were unable to continue their studies in the theory – initiative passed into the hands of the more resourceful British codebreakers (Grajek, 2019).

One might think that the British would be naturally inclined to continue their work more or less along the lines drawn by their Polish predecessors. Most of the young mathematicians entering the gates of Bletchley Park were educated in the intellectual tradition best epitomized by opinions by Godfrey Hardy, stressing the importance of pure vs applied mathematics (including well known "[r]eal mathematics has no effects on war. No one has yet discovered any warlike purpose to be served by the theory of numbers or relativity, and it seems very unlikely that anyone will do so for many years") (Hardy, 1940). In the reality of 1939/1940 attempts to continue algebraic attacks at the Enigma cipher would almost certainly lead to nowhere. So it was very fortunate that one of the first mathematicians to cross the gates of Bletchley Park was Alan Turing, who was never particularly concerned with the opinions of his professional circle and was usually following his own ways. This permitted him to create an interesting synthesis of the original, Polish algebraic approach with a new one, based on probability rather than algebra.

He took Rejewski's earliest discovery, that of Enigma cipher's cyclic property, as the foundation of his design; his machine was supposed to traverse the key space searching for the closed cycles (Turing, 1940). Contrary to Rejewski's original design he was not planning to search for these cycles within the message headers, but rather in the message contents. We might easily recognize Dilly Knox behind that decision. Immediately after his return from Pyry Knox expressed opinion that all Polish successes were based on a factor which might be removed by the adversary any moment: double encipherment of the message key. Dilly was right; that was precisely what happened on May 1st, 1940. At that time Turing bombe was already in the production process, and it did not rely on the analysis of the message indicator, so the change did not affect its construction.

There was, however, a price to be paid. Turing had designed his bomb so that it could search for the cycles within the fragment of the probable text (a crib) assumed by the codebreaker to be present in the coded message. His bombe was able to provide a solution only, and exclusively only, when this guess was right. Bombe's functioning was algebraic and deterministic with regard to the cycle search, and purely probabilistic with regard to the choice and position of the probable text. Later on Gordon Welchman strengthened the deterministic part of its job, adding the diagonal table, but overall the efficiency of the bombe was determined by the probabilistic component. As long as the codebreakers were able to provide a good and stable crib, they were able to break the key; otherwise the cipher remained invulnerable.

Bombe provided a practical solution for the networks of the German land and air army. Navy was using Enigma in much more ingenious way, resisting British attacks until late spring 1941. It was in the context of the struggle with the naval Enigma, that the British codebreakers switched entirely to the probabilistic attacks. Alan Turing and his colleagues proposed at least three different methods of attack at the naval Enigma, all of them based purely on the statistical properties of the cipher. E-rack represented most elementary of them, using the well-known fact that letter "e" represents most frequent letter in the German language, appearing in the written texts with outstanding frequency of nearly 17%. E-rack was based on the slightly paradoxical assumption that entire text being analysed consists of letter "e" only (Alexander, 1945). After the initial breakthrough with the naval Enigma E-rack assured several successes with the "Offizier" variant of the cipher.

Another method invented in the process was called "EINSing" (Alexander, 1945). Analysing decoded texts of German military messages the codebreakers have noticed that the most frequent single word encountered therein was EINS. They have designed a simple device enciphering EINS at each and every position of Enigma rotors and registering the result on the perforated cards. Then it was enough to register intercepted messages on the perforated cards and compare them, using the electromechanical sorter, with cards containing the EINS catalogue.

Third and most advanced method of attack on the naval Enigma was banburismus (Alexander, 1945). Its goal was to identify the right-hand Enigma rotor and, consequently, to narrow the number of rotor combinations being checked by the bombe. Banburismus represented the extension of the pre-war method proposed by Jerzy Różycki and referred to as the "clock method". Różycki used to analyse pairs of messages, whose indicators differed only in the last position; banburismus extended his approach for the pairs of messages differing in two, and under certain circumstances even thee positions. Codebreakers were registering incoming messages on the special sheets (manufactured in Banbury, hence the name of the method) and sliding pairs of sheets vs each other, looking for repeats. Every repeat one, two or three letters long was weighted using specially designed tables, measuring the probability of the coincidence. Sum of the partial results determined the probability that both messages were enciphered at the same or similar Enigma settings. It is worth noting that for the sake of banburismus Alan Turing invented the concept of "ban" – a measure of information equivalent to bit proposed by Claude Shannon. Banburismus was further extended to the "tetra catalogue" – repeats four or more letters long, processed using sorters and tabulators in the section called (from the name of its head) "Freebornery" (Alexander, 1945).

Neither of the described methods of attacks permitted breaking of the cipher directly. All of them were interdependent; efficient application of one depended on the earlier success of the other. Alan Turing and his colleagues had to wait until April/May 1941, when the documents captured onboard of some German ships permitted to overcome the crisis, and to start more or less regular operation of breaking the naval Enigma.

Their brief description above illustrates their nature sufficiently to recognize their purely probabilistic character. Under the pressure of the war necessity British codebreakers have given up algebraic approach, switching almost entirely to the well-established probabilistic and statistical methods. This tendency was further strengthened later on, during the attacks on the German teletype ciphers. Functioning of both devices constructed by the British codebreakers for this purpose, Heath Robinson and Colossus, relied on counting the measure of coincidence between the intercepted text and the pattern enciphered at every setting of the ciphering machine.

Two factors regarding British preference for probabilistic and statistical methods deserve additional comment. When Alan Turing was looking for a base for his banburismus, he decided to choose the less popular branch of statistics, the Bayesian inference, taking thus the position in the old debate between the *a priori* and *a posteriori* statisticians. Interestingly, using an *a priori* approach assured the German mathematicians about the security of Enigma ciphers (Ratcliff, 2003). Turing did not agree with a very principle of using an *a priori* approach; he argued that the ciphertext itself reveals some information about the cipher and the codebreaker should take this information into account. It was thus natural to reach for an *a posteriori* inference, and the Bayesian statistics provided a natural tool.

Second factor was of purely human nature. Most of the mathematicians recruited to Blechley Park belonged simultaneously to the top ranking group of chess players, at least in Britain, and some of them (among them C.H'O.D. Alexander) represented the top world level. Among the various circles, groups and clubs organized at the BP to provide recreation, chess club belonged to most numerous and most active.

So far no one was able to formulate an algebraic theory of the chess game; chess player naturally formulates his thinking about the game in terms of probability. It was thus natural to extend this model of thinking in the new game that the chess playing mathematicians were participating in.

As far as we know after the end of hostilities the codebreaking has for many years remained heavily dominated by the probabilistic and statistical methods. The landscape started to change only in late 1960s and early 1970s, when algebraic approach started to regain its citizenship rights in cryptology, being bravely accompanied by the number theory.

4 Algebraists & Probabilists

Although some simple statistical methods have been traditionally used in the codebreaking for over ten centuries, Polish success with Enigma in 1932 marked the real birth of the mathematical cryptology.

Interestingly, it was based on the oldest, perhaps next to geometry, field of mathematics, algebra. Rejewski's breakthrough was by no means accidental. Polish Cipher Bureau was the first cryptology agency in the world, which not only decided to employ mathematicians, but also expected, encouraged and trained them to apply their mathematical workshop in the codebreaking. Other codebreaking services followed its suit only after learning, directly or indirectly, about Polish success.

Methods of Enigma breaking invented by Polish team were somewhat exceptional. Their algebraic character made them deterministic: they granted breaking the cipher whenever Cipher Bureau was able to accumulate sufficient number of messages, without additional conditions regarding their contents. In that aspect they represented almost an antithesis of the then mainstream of traditional cryptanalysis, relying entirely on statistics and probability. Moreover, they were invented and used just in time to demonstrate their power. A few years later German crypto services started restructuring their operations, recruiting more mathematicians and permitting them to look at the codes, ciphers and machines from the perspective of their discipline. This new approach permitted to eliminate some mistakes and idiosyncrasies in the design of German ciphers, among them those, which made Polish approach so effective.

It was fortunate for the Allied cause that right at that moment the initiative in the attacks at Enigma ciphers passed into the British hands. The situation was developing in a somewhat paradoxical way. Marian Rejewski's studies in actuarial statistics indicate his interest in the applied mathematics. In spite of that he developed a theory of attack at the cipher in the best style of pure math. Most British codebreakers were educated in the respect for the pure math, and in spite of that decided to change the paradigm and switch to the probabilistic and statistical methods, the only ones practical considering the necessities of war and the only ones offering the prospects of success.

The history of attacks at Enigma ciphers is almost synonymous with the earliest period of the development of mathematical cryptology. It is fascinating to note that during that very early period Allied codebreakers developed and successfully applied methods based on two mutually complementary areas of modern cryptology; algebra on one part, and probability and statistics on the other. Present cryptanalysis relies on the mixture of both approaches. Its first stage usually involves the exploitation of cipher's algebraic properties to limit the search space. Then the probability and statistics take suit to find the solution within that limited space. It is interesting to note that precisely this approach provided the base for the construction of the Turing bombe.

References

Alexander C.H.O'D., *Cryptographic History of Work on German Naval Enigma*, 1945, NA HW 25/1

Al-Farahidi Ahmed, *Book of Cryptographic Messages*

Friedman, W.F., 1922, *The index of coincidence and its applications in cryptology*. Department of Ciphers. Publ 22. Geneva, Illinois, USA: Riverbank Laboratories

Grajek Marek, 2019, Sztafeta Enigmy. Odnaleziony raport polskich kryptologów, Agencja Bezpieczeństwa Wewnętrznego, Emów 2019

Hardy, G. H., 1940, *A Mathematician's Apology*, Cambridge University Press 1940

Al-Kindi, *Manuscript on Deciphering Cryptographic Messages*

Ratcliff R.A., 2003, *How Statistics Let the Germans to Believe Enigma Secure and Why They Were Wrong: Neglecting the Practical Mathematics of Cipher Machines*, Cryptologia 2003/2

Rejewski Marian, 1967, *Memories of my work at the Cipher Bureau of the General Staff Second*

Department 1930-1945, Adam Mickiewicz University Press, Poznań 2013

Turing Alan, 1940, *Prof's Book*, NA HW 25/3