The capture of the Enigma code machine 70 years ago changed the course of the second world war. But the secret codes broken by this event were not history's toughest ciphers. Plenty of codes are uncracked and, as MacGregor Campbell discovers, their solutions may provide the key to murders or even buried treasure.

**Killer codes**

**THE SOMERTON MAN**

On a warm summer's night in 1948, witnesses saw a well-dressed man lying on the beach in Somerton, near Adelaide, South Australia. By 6 a.m. the next morning, the man had not moved. He was dead.

An autopsy revealed organ damage consistent with poisoning, but no foreign substances in his body. He carried no identification. Fingerprint and dental-record searches came back with nothing. His clothes had no labels and were heavy for a night so balmy, suggesting he was not a local.

Thus was born the legend of the Somerton Man. It remains one of the most mysterious unsolved deaths in Australian history. To this day, we don't know who he was, a well-dressed drunk? A forlorn lover? A Russian spy? There is no shortage of theories, but facts are scarce. He did, however, leave behind a code.

Six months after the body was found, investigators discovered a small scrap of paper in a concealed pocket in his trousers. It read, simply, "Tamam Shud"—Persian for "ended". When investigators announced this find, it generated a new lead. Shortly after the death, a man who had packed his car and left it unlocked near the beach on the night Somerton Man was discovered found a copy of Persian poet Omar Khayyam's *Rubaiyat* in his car. The last page had a piece torn out of it that matched the mystery scrap in the Somerton Man's trousers. In the back of the book was a lettered code, scrambled over a few lines (pictured opposite).

Eventually all other leads pursued by the police came to dead ends, so this unsolved code is one of the few remaining clues that might yet reveal the man's identity.

Derek Abbott at the University of Adelaide is leading the most recent effort to untangle the mystery. His main approach uses software originally designed to identify the author of unknown texts. "My goal was not to crack the code, but to take a step back and apply statistical techniques to rule out which type of cipher techniques the "Tamam Shud" code is definitely not," he says.

He applied a simple "frequency analysis", a technique that credits counting the number of times each letter appears and that can spot the signatures of certain methods of coding. Though the brevity of the code makes it hard to be sure, Abbott and his team ruled out 20 different code types.

Abbott also claims to have ruled out the possibility that it is gibberish. His team had 30 people write random collections of letters and compared the results to the code. "The statistics of the letters show that the code bears structure and information content," he says.

His current theory is that the code is composed of the first letters of English words. The team is searching for patterns of words in e-books, looking for common passages that may have been used in the hidden message. They hope to soon expand their search to the entire web.

Abbott says that a faster route to figuring out the man's identity would be to exhume the body and analyse the DNA and the isotopes present in his bones. "The bone isotope test will reveal where he was born. With DNA we can find which family tree he belongs to." But unless he can persuade the police to take a fresh look at the case, the code is all we have to go on to solve this enduring mystery.

"To this day, we don't know who the man was. Well-dressed, drunk, forlorn lover or spy?"

A grisly find on Somerton beach (right), triggered a search for the identity of a dead man who left some baffling clues.

---

**BEALE'S BURIED TREASURE**

In 1881, James J. Beale published *The Beale Papers*, a pamphlet containing three coded messages. The solution of which, Ward claimed, would lead to sizeable buried treasure in the foothills of Sierra Nevada.

According to the pamphlet, 60 years earlier the enigmatic Beale left a small locked box containing the codes with an innkeeper before disappearing. The innkeeper passed the messages to a friend, who solved only one of the puzzles—the one containing the six riches awaiting whomever could solve the other two.

The solved code featured numbers that corresponded to the initial letters of the words in the US Declaration of Independence. The code was a string of numbers, to L. For example, "patriots" in the first letter of the twelfth word in the Founding Fathers' text.

In 1980, Jan Gillys used a computer to decode one of the remaining unsolved codes. He found a number of anomalies, such as long strings of characters in alphabetical order, suggesting they are a hoax.

There have been many subsequent efforts to find the treasure—and even a Hollywood movie called *National Treasure* inspired by the tale—but Gillys has seen nothing in the past 30 years to make him doubt his original assessment. "I'm quite confident that they're a hoax," he told New Scientist.
THE MIT PUZZLE

Nestled in the archives at the Massachusetts Institute of Technology is a metre-tall container made of lead. Its specific contents are a mystery, to be revealed only when an accompanying encrypted message is solved. Only one man has the solution: Brian Rivest, a professor of electrical engineering and inventor of the RSA algorithm, one of the most ubiquitous methods for encrypting online communications.

To celebrate his lab’s 37th anniversary in 1999, Rivest devised a “time-lock” puzzle inspired by his own encryption. Only when the solution is revealed, his rules state, may the lead bag be opened. He estimated it would take 35 years to solve—unless somebody could find a shortcut.

Rivest’s coded message is hidden in 616 numbers. The method of encryption differs from an alphabetic-based code, in which each letter matches another. It involves converting numbers into their binary form after solving a relatively difficult mathematical problem.

The problem that Rivest devised is simple to state: divide one number over 7.4 quadrillion digits long—approximately by a second number over 600 digits long. Then find the remainder.

The remainder, also over 600 digits long, is the “key” to unlocking the code. The codebreaker must convert it to binary form—a series of 0s and 1s—and compare it with a binary version of the 616 numbers of the original code. This comparison procedure is used to create a third string of binary: when the 1s and 0s match, that equals a 0; when they don’t, that’s a 1. Then you’re almost there: since binary can also represent the alphabet, the codebreaker would translate this third string of binary into the letters of the secret message.

So why would that take 35 years to solve? Rivest designed the message to be so big that it would take decades of continuous computation to calculate—it’s hard to assume that computing speed doubles roughly every two years, as predicted by Moore’s law. The calculation can only be done in sequential steps, says Rivest, making it impossible to distribute computer networks or the parallel processors used in supercomputers.

“Let’s say Rivest had designed the puzzle to be a 20 minute problem,” Prof. John Moeller of the University of California, San Diego, told the New Scientist. “There would still be a shortcut. The second 600-digit number in the calculation is the product of two large prime numbers. If those could be found, the computation could be completed in much less time, because a different mathematical technique could be applied.”

Finding these factors would be no mean feat, however. The difficulty of factoring products of two large primes is the core mathematical fact underlying RSA encryption, a widespread method to allow two parties online to communicate securely without both of them needing a secret key. Rivest developed it in 1977 along with Adi Shamir and Len Adleman.

For example, if Alice wished to send Bob a secret message, she could ask him for a number and do a relatively simple and non-reversible calculation to encrypt the message. That number, the message’s “public key” is not secret. It is essentially the product of two large prime numbers. Only Bob can decrypt the message because only he knows what the two primes are: the largest public key number ever “broken” into its two primes was achieved by a team led by Thorsten Kleinjung at the Swiss Federal Institute of Technology in Lausanne. It was 135 digits long, so it wasn’t close to the size of the 600-plus-digit number in Rivest’s puzzle.

Rivest told New Scientist that he thinks his original estimate of how long the puzzle would take may be too optimistic. “Computing power isn’t increasing quite as fast as predicted, in terms of ability to do fast sequential computations,” he says. “Evaluating the cost and guessing the message will likely take longer than expected.

Still, Rivest has underestimated codebreakers in the past. In 1977, he helped to design a puzzle for Martin Gardner’s recreational mathematics column in Scientific American. The solution to the puzzle required factoring a 129-digit number. Nobody managed it at the time, but in 1979 a team of 600 volunteers did it in eight months, helping to reveal the message: “The Magic Words Are Squeamish Ossifrage.” (Ossifrage is the name of a type of vulture, and means “bone-breaker.”) Rivest and colleagues had estimated, based on the computing power of the time, that the task would take 40 quadrillion years.

KRYPTOS

At the U.S. Central Intelligence Agency’s headquarters in Langley, Virginia, there is a monument to secrecy that has mixed professional and amateur codebreaker for over two decades. Erected in 1959, the Kryptos sculpture is a corporate gift from the RAND Corporation. Its creator is American artist James Sanborn, who received cryptography training from a CIA code expert. The sculpture’s code is divided into four sections. Three of these sections are now broken, revealing epigraphic messages that hint at a deeper mystery. But the fourth and final section has yet to be cracked.

Sanborn has hinted that the clues to unlocking it lie in the first three sections, which a Californian computer scientist named Jim Gillogly announced he had solved in 1999. Shortly after, the CIA announced that some of their analysts, David Stein, had solved the same sections in 1999, with pencil and paper.

The first two sections were encoded using a modified form of a Vigenere cipher. Third, an American government-grade letter substitution technique, is used for each of the 26 different columns. As with the Vigenere cipher, the key is a sequence of 26 different keys in an appropriate order. For example, to encrypt the message “HELLO WORLD”, the encryption process would be:

```
HELLO
WORLD
```

This message is then divided into 26 parts, each of which is encrypted using a different key. The result is a long string of letters and numbers that cannot be easily deciphered without knowing the key.

Rivest is also the first codebreaker to find a shortcut to the Kryptos puzzle. His method involves analyzing the ciphertext and looking for patterns that might indicate how the encryption was done. By doing this, Rivest has been able to make some interesting observations about the structure of the ciphertext and how it was created. For example, he has noticed that the ciphertext contains a number of repeating patterns that suggest that the encryption was done using a key with a very large number of symbols. This is consistent with the idea that the key is a long string of letters and numbers that cannot be easily deciphered without knowing the key.

Cracking a Vigenere cipher

A Vigenere cipher can be cracked using a grid of multiple alphabets called a tabula recta. If the codebreaker can find the keyword, it will help them to read the columns of the grid and translate the message into the hidden message.

Step 1: The code

Faced with this, the codebreaker must first use trial and error, computer analysis and other clues to discover a secret keyword.

Step 2: The key

In this case the keyword is “KEY”. The codebreaker locates this word in the grid and uses it to decipher each letter of the ciphertext.

Step 3: The route

For the first coded letter, R, the codebreaker can use the X column and K to find the matching letter from the grid. Reading across the row reveals that the first letter of the hidden message is H.

Step 4: The solution

To decode the next letter of the message, the codebreaker turns to the E column, then reads down to find the matching letter. Repeating this process until the full message is revealed.

```
HIDDEN MESSAGE
```

```
CRACK THIS: The baby boomer founded the code-breaking firm Hunt B Enterprises. In 1998, he sold his shares in Kryptos to the CIA. His company was called “Computer Solution of Mysteries.”
```

```
25 May 2011 | New Scientist | 43
```
THE VOYNICH MANUSCRIPT

As incongruous as go, few inspire as
many imaginative interpretations as the
Voynich manuscript, a medieval tome
filled with illustrations of medicinal
plants, astrological diagrams, naked
nymphs and pages of indecipherable
script. Despite a century of effort by
some of the greatest cryptanalysts,
not a word has been decoded.

Little is known about the book’s
history prior to 1912, when book dealer
Wilfrid Voynich found it in an Italian
monastery, though it is believed to have
belonged to the Holy Roman Emperor
Rudolph II of Bohemia.

In 2004, computer scientist and
linguist Gordon Rugg at the University
of Keele, UK, published a persuasive
argument that the manuscript is, in
fact, nonsensical. This contradicted
previous researchers, who had noted
that the pattern of word lengths and
symbol combinations was similar
to structures found in real languages,
a match that would be difficult for
a forger to replicate. But Rugg showed
that it would be possible to create
a nonsense similar to that found in
the manuscript using techniques
known to 15th-century cryptographers
(Cryptologia, vol 28, p 31).

Andreas Schöner at Johannes Kepler
University in Linz, Austria, published a
study supporting this interpretation in
2007. He found that the statistical
properties of the script are consistent
with gibberish (Cryptologia, vol 31, p 95).

It is completely new, however. In 2009,
radiotherapists dating pinned down the age
of the book’s vellum to the mid-15th
century. If the age was more recent,
a false would be certain.

Schöner says this does not rule out a
heist, but concludes that it is not possible
to prove mathematically that the last
doctor or does not hold meaning.

CODES OF THE SECOND WORLD WAR

The second world war brought a shift from
handwritten cryptography to machines
capable of spitting up complex, ever-changing
codes. The best known of these was the
Enigma machine, first used by the German
navy in 1918. Enigma machines used three or
four mechanical rotors to scramble electrical
circuits that assigned the letters of the
message to be encrypted into ciphers.
The rotors were changed regularly, often every
day, meaning that messages seldom shared the
same encryption.

With the help of Alan Turing, and early
cryptanalysts called cryptologists, Allied
cryptoanalysts eventually cracked many
Enigma messages, which helped to turn the tide of the war. Many
messages, however, were never broken.

In 2006, Stephen Kuhn, an amateur
cryptoanalyst based in Utrecht, the
Netherlands, began an effort to crack three
such Enigma messages intercepted by HMS
Hurricane in the north Atlantic in 1942. Like
SETI Lighthouse, which seeks to harness space
home-computer capacity to analyze signals
from space for signs of intelligent life, Kuhn’s
EnigmaLighthouse harnesses thousands of
computers around the world to help crack
the codes. The project solved the first two
messages in a matter of months—the submarine
location report and how a crew was forced
to surrender after coming under attack. The
third has yet to be cracked.

David Kahn, author of The Code breakers,
so an offshoot history of cryptography, says it’s
not surprising that an Enigma message has
not been cracked such a large computational
attack. The Allies only broke Enigma transmissions
by taking educated guesses at the solution,
then using Babbage to generate possible settings
that yielded the decoded text. Without an
educated guess as to the content, he says,
there are too many combinations of wheel
to settings to break Enigma in a human lifetime.

THE DORABELLA CODE

The English composer Edward Elgar was a keen
cryptographer. The melody of his Enigma
Variations—which for which the German
encoding machines were named—is supposedly
complementary to the melody of a famous
song by another composer. He didn’t say which.

This melody is the only surviving
Elgar puzzle. In 1897, he wrote an 87-character
code to his friend Dorabella Penny. Forty years
later, she published the code in her memoirs but
clicked never to have solved it.

In the intervening years, many would be
codebreakers have also been stumped. The script
appears to contain 24 distinct squiggly symbols
spread across three lines. Analysis of the code
suggests the symbols could be a simple
"substitution cipher," where each symbol is
assigned to a letter. But this hasn’t produced an
answer, so perhaps it is a shorthand language
shared only between Elgar and Penny.

In 2007, the Elgar Society organised a
competition to break the code. Alex, none
of the entries was convincing.

THE ZODIAC KILLER

Famed for peace, love, and tie-dye, San Francisco
in the late 1960s was also the setting for one of the creepiest
real-life murder mysteries in American history. From
December 1968 to October 1969, a serial killer using the
moniker "Zodiac" murdered at least seven people. The killer
taunted local police and newspapers with handwritten
t Rewrite questions
letters containing threats, claims of further
uncovered victims, and coded messages.

The supposed killer claimed
that the solutions would reveal
an identity when they were all
solved. Some never came, and
the killer was never caught.

The first three codes were
encrypted by replacing letters
with symbols. But there was a
trick: at the core of the
most frequent letters, like "e" and "t,"
was assigned to multiple symbols.

This made it easier to solve
by simple codebreaking
techniques like looking for
more common symbols out of
the characters in the rows
should be equal to the distributions
of ones in the columns, but that
isn’t the case.

In 2009, Ryan Garthc, a
student at the University of North
Texas in Denton, along with his
students, tried to "evolve"
a solution using a so-called
Genetic algorithms. They first
randomly generated possible
pairings of English letters with
symbols from the code. They
then looked for how well a
given set of pairings of Zodiac
symbols and English letters
produced two-three-letter sequence in the potential
solution, like "th" and "es," which are common English.
The best performing
pairings were subjected to further selection
and recombination. Using this method, they were able
to evolve the solution to the
first, second, Zodiac message, but made little progress with
the third. The code
mixing and mutation
by researchers at San Jose State University cryptography
have yielded similar
disappearing results.

Garthc suggests that some sort of
physical rearrangement of the symbols
is necessary. Testing all the possible ways this could be done,
however, is daunting. "You have to happen upon
exactly the right thing before any of our software tools
work," he says.

MacGregor Campbell is a
consultant for New Scientist.