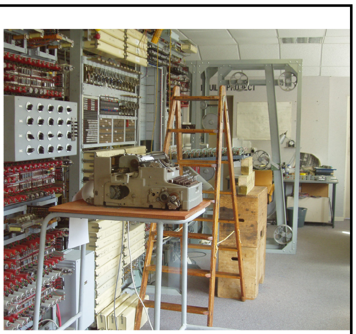
Lecture 17: Double Deltas and Banburismus



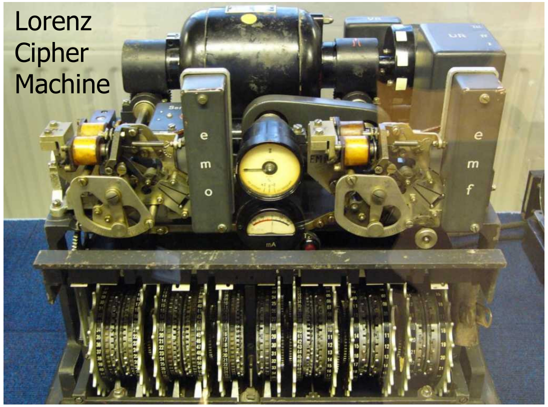
Colossus Rebuilt, Bletchley Park, Summer 2004

We'll finish up the tree sorting next week

Today's notes: on-line only (links)

CS150: Computer Science
University of Virginia
Computer Science

David Evans
http://www.cs.virginia.edu/evans



Lorenz Cipher Machine

Lorenz Wheels

12 wheels 501 pins total (set to control wheels)

Work to break in $\Theta(p^w)$ so real Lorenz is $41^{12}/5^3$ ~ 1 quintillion (10^{18}) times harder!

Breaking Fish

- GCHQ learned about first Fish link (Tunny) in May 1941
  - Intercepted unencrypted Baudot-encoded test messages
- August 30, 1941: Big Break!
  - Operator retransmits failed message with same starting configuration
  - Gets lazy and uses some abbreviations, makes some mistakes
    - SPRUCHNUMMER/SPRUCHNR (Serial Number)

"Two Time" Pad

- Allies have intercepted:

  $C1 = M1 \oplus \mathbf{K1}$

  $C2 = M2 \oplus \mathbf{K1}$

  Same key used for both (same starting configuration)
- Breaking message:

$$\begin{aligned} C1 \oplus C2 &= (M1 \oplus \mathbf{K1}) \oplus (M2 \oplus \mathbf{K1}) \\ &= (M1 \oplus M2) \oplus (\mathbf{K1} \oplus \mathbf{K1}) \\ &= M1 \oplus M2 \end{aligned}$$

"Cribs"

- Know: C1, C2 (intercepted ciphertext)

  $C1 \oplus C2 = M1 \oplus M2$
- Don't know M1 or M2
  - But, can make some guesses (cribs)
    - SPRUCHNUMMER
    - Sometimes allies moved ships, sent out bombers to help the cryptographers get good cribs
- Given guess for M1, calculate M2

  $M2 = C1 \oplus C2 \oplus M1$
- Once guesses that work for M1 and M2

  $K1 = M1 \oplus C1 = M2 \oplus C2$

Reverse Engineering Lorenz

- From the 2 intercepted messages, Col. John Tiltman worked on guessing cribs to find M1 and M2: 4000 letter messages, found 4000 letter key K1
- Bill Tutte (recent Chemistry graduate) given task of determining machine structure
  - Already knew it was 2 sets of 5 wheels and 2 wheels of unknown function
  - Six months later new machine structure likely to generate K1

Intercepting Traffic

- Set up listening post to intercept traffic from 12 Lorenz (Fish) links
  - Different links between conquered capitals
  - Slightly different coding procedures, and different configurations
- 600 people worked on intercepting traffic

Breaking Traffic

- Knew machine structure, but a different initial configuration was used for each message
- Need to determine wheel setting:
  - Initial position of each of the 12 wheels
  - 1271 possible starting positions
  - Needed to try them fast enough to decrypt message while it was still strategically valuable

This is what you did for PS4 (except with fewer wheels)

Recognizing a Good Guess

- Intercepted Message (divided into 5 channels for each Baudot code bit)

  $Z_c = z_0 z_1 z_2 z_3 z_4 z_5 z_6 z_7 \ldots$

  $z_{c,\,i} = m_{c,i} \oplus x_{c,i} \oplus s_{c,i}$

  Message Key (parts from S-wheels and rest)
- Look for statistical properties
  - How many of the $z_{c,i}$'s are 0? ½ (not useful)
  - How many of $(z_{c,i+1} \oplus z_{c,i})$ are 0? ½

Double Delta

$$\Delta Z_{c,i} = Z_{c,i} \oplus Z_{c,i+1}$$

Combine two channels:

$\Delta Z_{1,i} \oplus \Delta Z_{2,i} = \Delta M_{1,i} \oplus \Delta M_{2,i}$ > ½ Yippee!

$\oplus \Delta X_{1,i} \oplus \Delta X_{2,i}$ = ½ (key)

$\oplus \Delta S_{1,i} \oplus \Delta S_{2,i}$ > ½ Yippee!

Why is $\Delta M_{1,i} \oplus \Delta M_{2,i} > ½$

Message is in German, more likely following letter is a repetition than random

Why is $\Delta S_{1,i} \oplus \Delta S_{2,i} > ½$

S-wheels only turn when M-wheel is 1

Actual Advantage

- Probability of repeating letters

  $\text{Prob}[\Delta M_{1,i} \oplus \Delta M_{2,i} = 0]$ ~ 0.614

  3.3% of German digraphs are repeating
- Probability of repeating S-keys

  $\text{Prob}[\Delta S_{1,i} \oplus \Delta S_{2,i} = 0]$ ~ 0.73

$\text{Prob}[\Delta Z_{1,i} \oplus \Delta Z_{2,i} \oplus \Delta X_{1,i} \oplus \Delta X_{2,i} = 0]$

$= 0.614 * 0.73 \quad + (1-0.614) * (1-0.73)$

Δ M and S are 0 Δ M and S are 1

= 0.55 **if** the wheel settings guess is correct (0.5 otherwise)

Using the Advantage

- If the guess of **X** is correct, should see higher than ½ of the double deltas are 0
- Try guessing different configurations to find highest number of 0 double deltas
- Problem:

  # of double delta operations to try one config

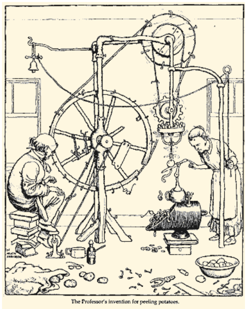
  = length of Z * length of X

  = for 10,000 letter message = 12 M for each setting * 7 ⊕ per double delta

  = 89 M ⊕ operations

Heath Robinson

- Dec 1942: Decide to build a machine to do these ⊕s quickly, due June 1943
- Apr 1943: first Heath Robinson machine is delivered!
- Intercepted ciphertext on tape:
  - 2000 characters per second (12 miles per hour)
  - Needed to perform 7 ⊕ operations each ½ ms

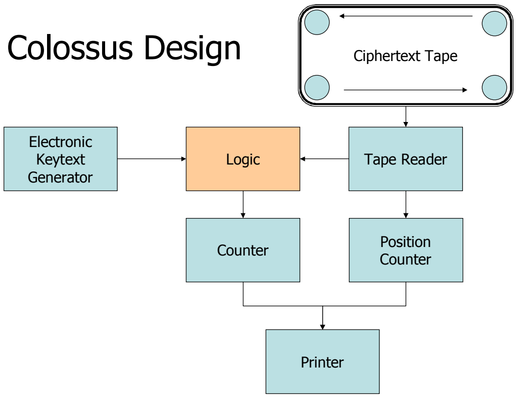


Heath Robinson, British Cartoonist (1872-1944)

Colossus

- Heath Robinson machines were too slow
- Colossus designed and first built in Jan 1944
- Replaced keytext tape loop with electronic keytext generator
- Speed up ciphertext tape:
  - 5,000 chars per second = 30 mph
  - Perform 5 double deltas simultaneously
  - Speedup = 2.5X for faster tape * 5X for parallelism

Colossus Design



Ciphertext Tape

Electronic Keytext Generator

Logic

Tape Reader

Counter

Position Counter

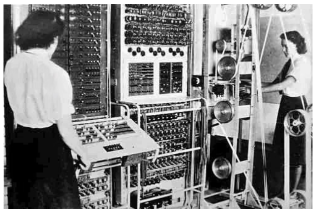
Printer

Impact on WWII

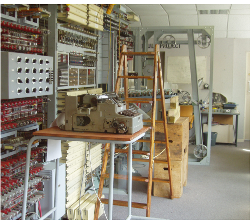
- 10 Colossus machines operated at Bletchley park
  - Various improvements in speed
- Decoded 63 million letters in Nazi command messages
- Learned German troop locations to plan D-Day (knew the deception was working)

Colossus History

Kept secret after the war, all machines destroyed

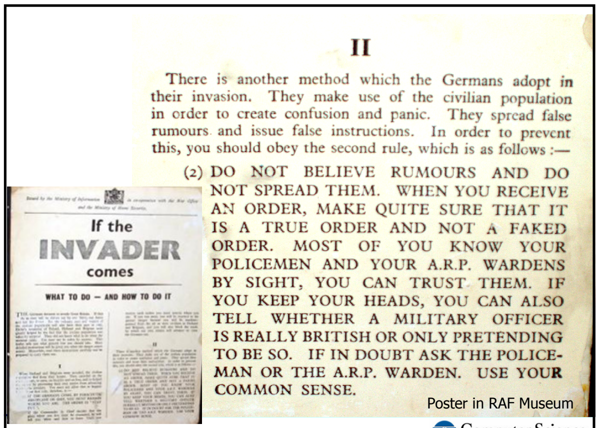


During WWII



Rebuild, Bletchley Park, Summer 2004

How could the folks at Bletchley Park solve a problem ~ 1 quintillion times harder than ps4?



If the INVADER comes

WHAT TO DO — AND HOW TO DO IT

II

There is another method which the Germans adopt in their invasion. They make use of the civilian population in order to create confusion and panic. They spread false rumours and issue false instructions. In order to prevent this, you should obey the second rule, which is as follows :—

(2) DO NOT BELIEVE RUMOURS AND DO NOT SPREAD THEM. WHEN YOU RECEIVE AN ORDER, MAKE QUITE SURE THAT IT IS A TRUE ORDER AND NOT A FAKED ORDER. MOST OF YOU KNOW YOUR POLICEMEN AND YOUR A.R.P. WARDENS BY SIGHT, YOU CAN TRUST THEM. IF YOU KEEP YOUR HEADS, YOU CAN ALSO TELL WHETHER A MILITARY OFFICER IS REALLY BRITISH OR ONLY PRETENDING TO BE SO. IF IN DOUBT ASK THE POLICEMAN OR THE A.R.P. WARDEN. USE YOUR COMMON SENSE.

Poster in RAF Museum

Motivation Helps…

Confronted with the prospect of defeat, the Allied cryptanalysts had worked night and day to penetrate German ciphers. It would appear that fear was the main driving force, and that adversity is one of the foundations of successful codebreaking.

Simon Singh, *The Code Book*

The Good News...

No problems on your exam are 1/quintillionth as hard as breaking the Lorenz cipher

Banburismus

Given two Enigma-encrypted messages, how can we determine if they were encrypted starting with the same wheel settings?

Enigma

Enigma machine at Bletchley Park

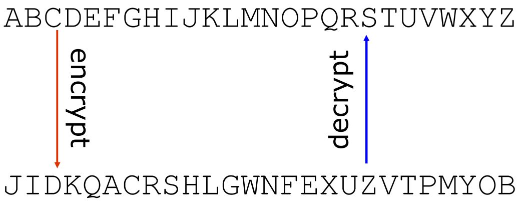
- Invented commercially, 1923
- German Navy, Army, Air Force
- About 50,000 in use (many were captured by Allies)
- Modified throughout WWII, Germans believed perfectly secure
- Kahn's *Codebreakers* (1967) didn't know it was broken
- Turing's 1940 Treatise on Enigma declassified in 1996

Reverse Engineering Enigma

"This fictional movie about a fictional U.S. submarine mission is followed by a mention in the end credits of those actual British missions. Oh, the British deciphered the Enigma code, too. Come to think of it, they pretty much did everything in real life that the Americans do in this movie."

Roger Ebert's review of **U-571**

Simple Substitution Ciphers



CS ⇒ DZ

Rotor Wheels

Simple substitution

Latch turns next rotor once per rotation

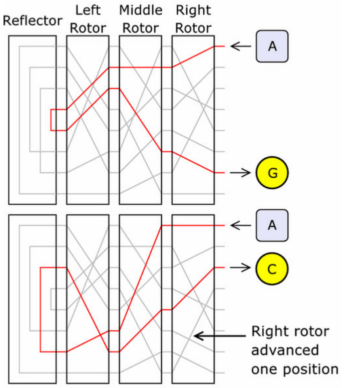
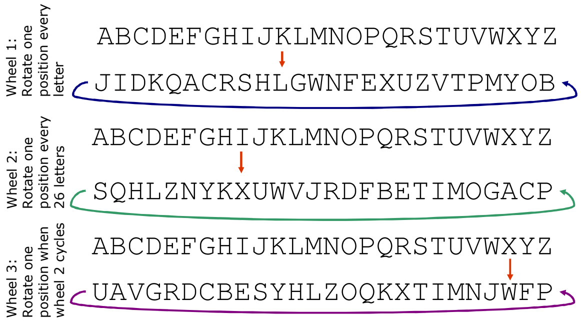


Image from http://en.wikipedia.org/wiki/Image:Enigma-action.png

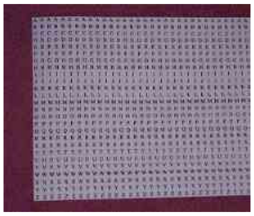
Enigma's Rotating Substitutions


Wheel 1: Rotate one position every letter
ABCDEFGHIJKLMNOPQRSTUVWXYZ
JIDKQACRSHLGWNFEXUZVTPMYOB
Wheel 2: Rotate one position every 26 letters
ABCDEFGHIJKLMNOPQRSTUVWXYZ
SQHLZNYKXUWVJRDFBETIMOGACP
Wheel 3: Rotate one position when wheel 2 cycles
ABCDEFGHIJKLMNOPQRSTUVWXYZ
UAVGRDCBESYHLZOQKXTIMNJWFP

Language is Non-Random

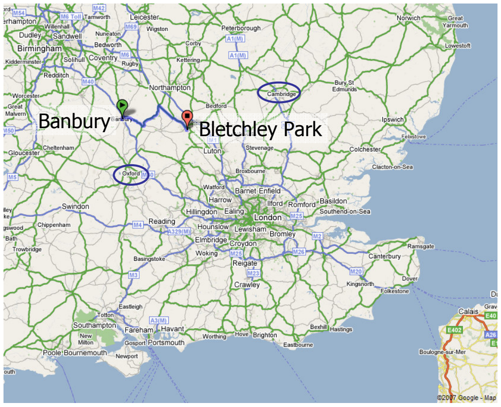
- Random strings: the probability of two letters in the two messages matching is 1/26 (number of letters in alphabet)
- Same-encrypted strings: the output letters will match when the input letters match
  - This happens much more frequently because some letters (e.g., "e" is ~13% of all letters) are more common

Alan Turing's Solution

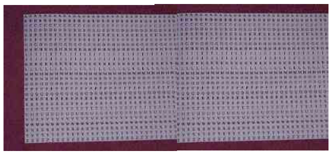


M1: GXCYBGDSLVWBDJLKWIPEHVYGQZWDTHRQXIKEESQS

M2: YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCFQAXVXDVU



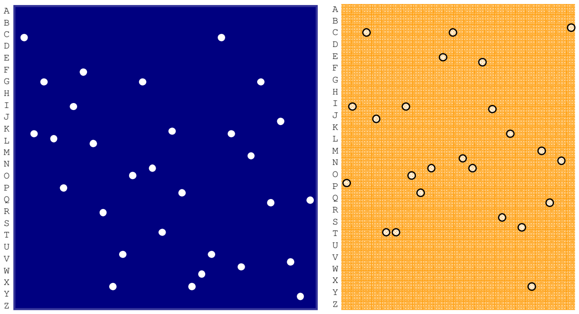
Banburismus



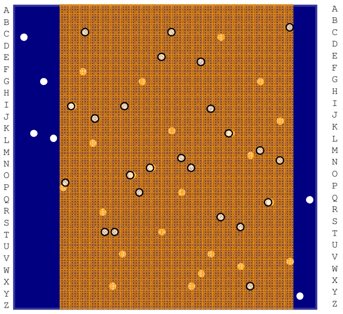
M1: GXCYBGDSLVWBDJLKWIPEHVYGQZWDTHRQXIKEESQS

M2: YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCFQAXVXDVU

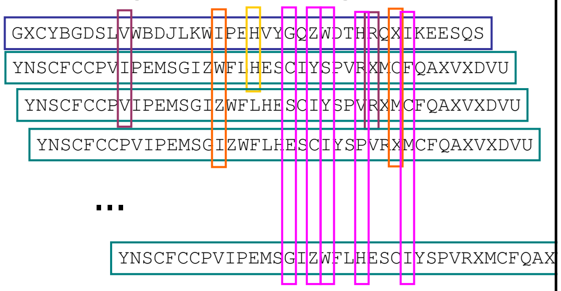
Intercepted Message 1 Intercepted Message 2


A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Intercepted Message 1


A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Trying Possible Alignments



GXCYBGDSLVWBDJLKWIPEHVYGQZWDTHRQXIKEESQS

YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCFQAXVXDVU

YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCFQAXVXDVU

YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCFQAXVXDVU

…

YNSCFCCPVIPEMSGIZWFLHESCIYSPVRXMCFQAX

Turing’s Hut 8 at Bletchley Park

Don’t complain about your working space (or Small Hall).
You can do good computer science anywhere.
But find a quiet, undisturbed place to work on the exam.

Charge

- Exam 1: Out now, due beginning of class Monday (2 minute grace period)
- Please please please be honorable!
  - I want the other exams to be take-home also
- Last 2 questions are about Banburismus

- When in London, visit Bletchley Park (about 1 hour train ride)