

Ganzúa: A Cryptanalysis Tool for Monoalphabetic and Polyalphabetic Ciphers

JESÚS ADOLFO GARCÍA-PASQUEL and JOSÉ GALAVIZ

Universidad Nacional Autónoma de México

Many introductory courses to cryptology and computer security start with or include a discussion of classical ciphers that usually contemplates some cryptanalysis techniques used to break them. Ganzúa (picklock in Spanish) is an application designed to assist the cryptanalysis of ciphertext obtained with monoalphabetic or polyalphabetic ciphers. It can use almost arbitrary character sets for the plain and cipher alphabets as well as obtain the standard relative frequencies of many languages and provide other useful data.

Categories and Subject Descriptors: E.3 [**Data Encryption**]: Code breaking; K.3.1 [**Computers and Education**]: Computer Uses in Education—*Computer-assisted instruction (CAI)*; K.3.2 [**Computers and Education**]: Computer and Information Science Education—*Computer science education*; K.8.1 [**Personal Computing**]: Application Packages—*Freeware / shareware*

General Terms: Security

Additional Key Words and Phrases: Cryptology, classical cryptography

1. INTRODUCTION

A clear understanding of cryptology is essential to the study of computer security. This is the reason why most introductory courses spend a considerable amount of time discussing classical ciphers [Cummings; Quer 1996; Ramió 2003; Weibel 2002]. These cryptographic systems can illustrate many concepts and principles. The cryptograms they generate can also be simple enough to be solved relatively quickly, which gives the students an opportunity to familiarize themselves with different kinds of cryptanalytic attacks and techniques, as well as understand Kerckhoffs' maxims [Kerckhoffs 1883], particularly, why "*only a cryptanalyst can judge the security of a crypto system.*" This way we provide a solid background for the study of modern symmetric cryptosystems. As easy to solve as the cryptograms obtained using classical cyphers may be, gathering

Author's address: José Galaviz, Cubículo 019, Departamento de Matemáticas, Facultad de Ciencias, UNAM. Circuito Exterior, Ciudad Universitaria, 04510, México D.F., México. email: adolfo@ciencias.unam.mx; jgc@ciencias.unam.mx

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or direct commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.
© 2006 ACM 1531-4278/06/0900-0001 \$5.00

the information required for the cryptanalysis (or writing programs to do it) is a time-consuming process that distracts from the main objectives of the exercise.

Ganzúa [Garcia-Pasquel 2004] is a cryptanalysis tool for monoalphabetic and polyalphabetic ciphers that simplifies the cryptanalysis process by obtaining the required information. It was developed to assist an introductory Cryptology course at the Facultad de Ciencias (Faculty of Science) of the UNAM that is targeted at undergraduate computer science and mathematics students [Galaviz and Magidin 2003]. The application is open source software and can be found at <http://ganzua.sourceforge.net>

There are some programs similar to Ganzúa available on the Internet [Bobbit; Peschel; Russell; Spillman 2002], but they share some characteristics that make them inconvenient for the teaching of classic cryptanalysis, especially in languages other than English. Most of these applications arbitrarily limit the characters that may be used in the ciphertext to the 26 letters of the English alphabet (not even the set of characters in the International Morse Code), so they can not be used effectively with cryptograms that include other characters (or more than 26). Beyond the limitation posed by the set of characters and its size, many can only be used to solve monoalphabetic ciphers. Since the course is given in the Spanish language and uses both monoalphabetic and polyalphabetic ciphers on its first part, a decision was made to develop a tool for these cryptographic systems and to make it as flexible as possible, so it could be used with arbitrary character sets and languages. Using a particular language or set of characters does not change the concepts at the core, but can help the students analyze the cryptograms.

The following are some of Ganzúa's qualities:

- Features present for all ciphers:
 - Support for arbitrary plain and cipher alphabets with characters chosen among those in the Unicode Standard character set.
 - Obtaining and display of the standard relative frequencies of characters, digrams, and trigrams of languages.
 - Index of coincidence computation for cryptograms and languages.
 - Estimation of the number of alphabets used to generate a cryptogram, based on the index of coincidence of the ciphertext and the language.
 - Enforcement of injective character substitution.
 - Cryptanalysis project saving and loading.
- Monoalphabetic cryptanalysis features:
 - Alphabet-wide substitution tools for the Caesar shift cipher and other monoalphabetic ciphers.
 - Obtaining and display of the relative frequencies of characters, digrams, and trigrams in cryptograms.
- Polyalphabetic cryptanalysis features:
 - Alphabet-wide substitution tools for the Vigenère or Alberti ciphers.
 - Obtaining and display of the relative frequencies of characters ciphered using each alphabet.
 - Kasiski Test exertion.

Ganzúa is a Java application, so it is platform-independent. It was important that the program could be executed on many platforms, because the students have access to computers with different architectures and operating systems.

2. MONOALPHABETIC SYSTEMS

The monoalphabetic ciphers are useful to introduce many concepts, like symmetric cryptography, ciphertext and plaintext. They can also be used to illustrate some principles, for example, that using a long series of rules (like applying repeatedly a set of monoalphabetic substitutions) may be just an illusive complication [Sinkov 1990].

Many students at the course have already been familiarized with the Caesar cipher through cereal boxes or the use of ROT13 in online forums, but most of them have never analyzed a cryptogram generated with a mixed alphabet. This gives us the opportunity to introduce some notions of cryptanalysis as well as the frequency and contact analysis techniques.

Ganzúa can assist the students by providing the relative frequencies of characters, digrams, and trigrams in the ciphertext, as well as the standard frequencies of the language, sorting them, and making character substitution a matter of selecting the substitution from a list. This keeps the students focused on the interpretation of the data and makes the cryptanalysis process much faster. Figures 1 and 2 show the data displayed by Ganzúa while solving a cryptogram obtained by applying a monoalphabetic cipher to some text¹ in English. While the standard relative frequencies displayed in this case belong to the English language, Ganzúa also includes those of Spanish and a tool to obtain them for other languages and alphabets.

Although the course uses mixed alphabets in most of its examples of monoalphabetic ciphers, Ganzúa includes tools to manipulate the substitution (e.g. reverse or shift it) that can be useful for other systems like Caesar or Atbash [Kahn 1999]. This way, if the students suspect that an exercise was obtained using one of these ciphers after looking at the relative frequencies, they can attempt to guess with a few clicks.

3. POLYALPHABETIC SYSTEMS

The polyalphabetic systems provide some of the best examples of why the aforementioned statement quoted from Kerckhoffs is true and allow us to use the first analytic test in the history of cryptanalysis, Friedman's Index of Coincidence [Bauer 2000; Kahn 1999; Sinkov 1990].

To begin the study of these systems, we introduce the Vigenère cipher [Kahn 1999] with its usual square and point out that each letter of the keyword determines a Caesar cipher to be used. Then, after an example or two using the traditional square, we mention that Vigenère's original method did not specify a unique table, but that it could have mixed alphabets on its headers and would rotate yet another inside it. To avoid any confusion between the two, we call the system that uses the traditional table Vigenère cipher, and the one with

¹In this case a fragment of Arthur Conan Doyle's "*The Great Boer War*"

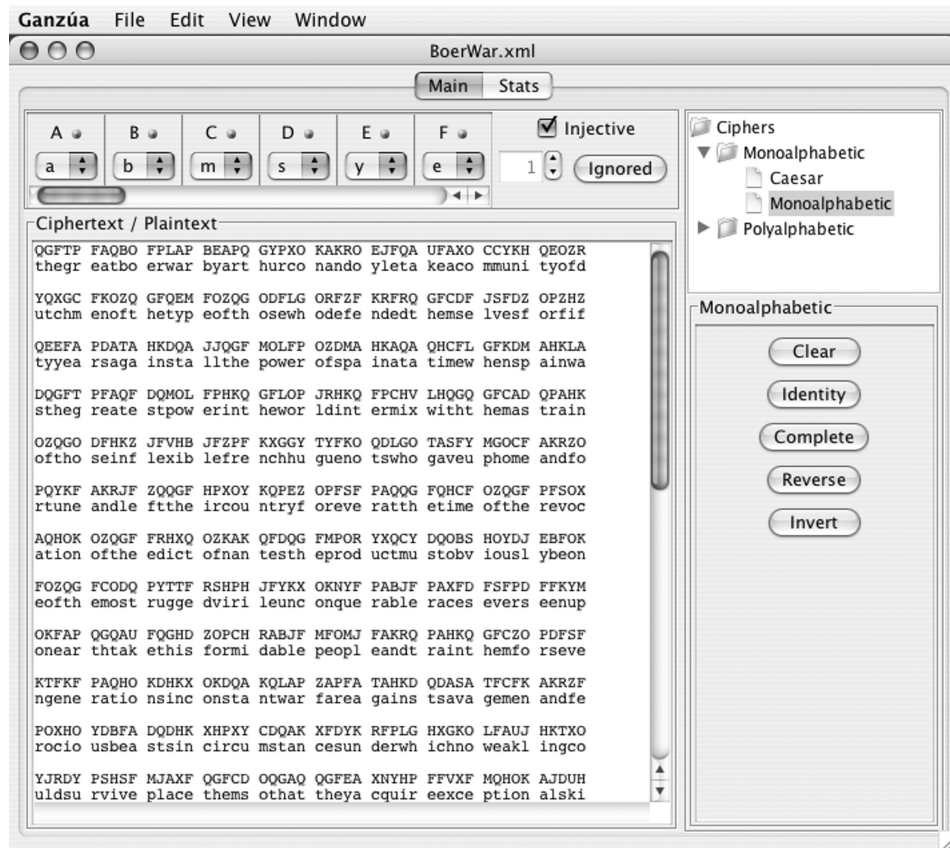


Fig. 1. Ganzúa's main window.

arbitrary tables Alberti cipher [Kahn 1999]. At this point, the Babbage–Kasiski Test and Friedman's Index of Coincidence are explained [Bauer 2000; Sinkov 1990].

Now that the students know how to cipher and some cryptanalysis techniques, they can benefit from what Ganzúa has to offer. Ganzúa can perform the Babbage–Kasiski Test and find all the repeated n -grams in the ciphertext, the distance between consecutive occurrences and the prime factors of those distances as well as sort the sequences by length or frequency in the text (see Figure 3). It also calculates the Index of Coincidence and can give an estimate of the number of alphabets used (as seen in Figure 4).

We begin by solving examples that use the Vigenère cipher. Once the length of the keyword is identified using the Kasiski and Friedman tests, Ganzúa can get the relative frequencies of the characters that were ciphered using each alphabet. By comparing these frequencies with those of the language, we can identify each of the alphabets with relative ease, since all of the alphabets are just rotations. In addition, using the *identity* and *shift* buttons along with the statistical data, a cryptogram ciphered with the Vigenère system can be solved in a matter of minutes.

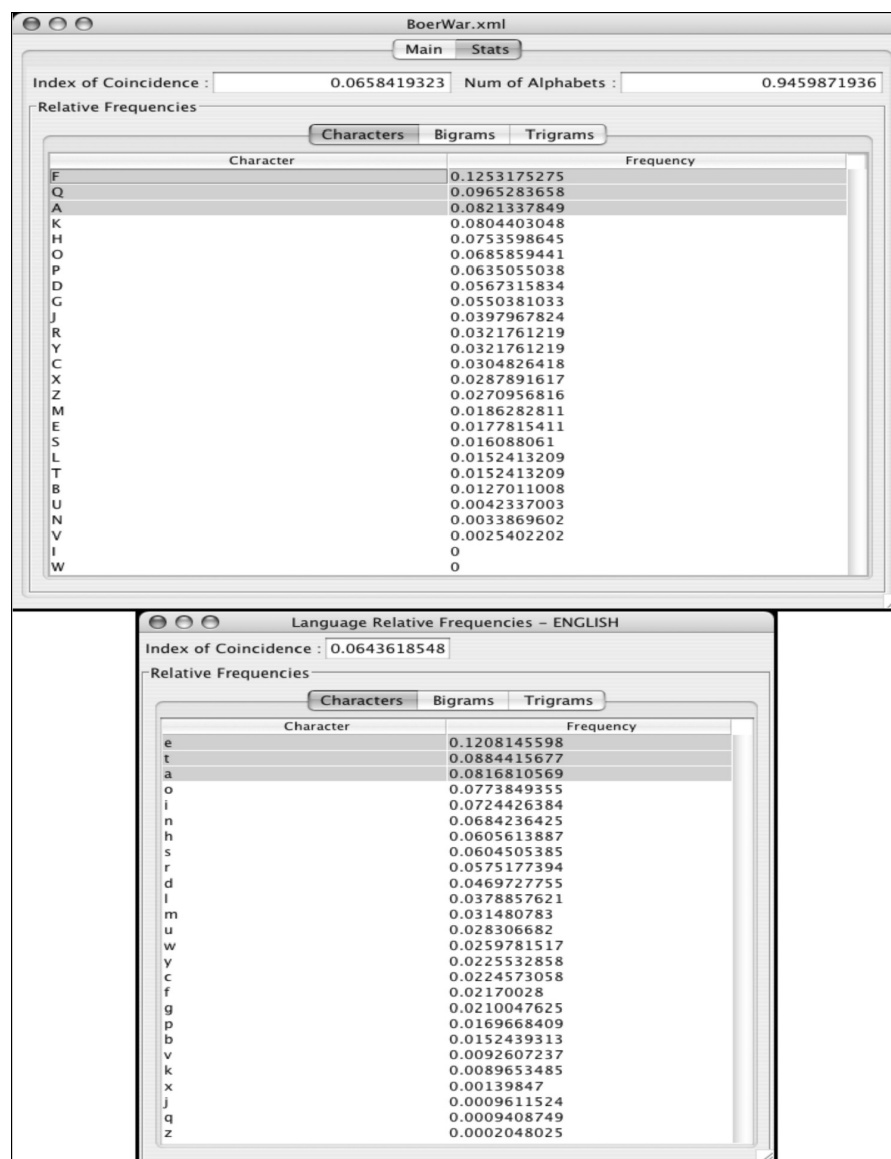
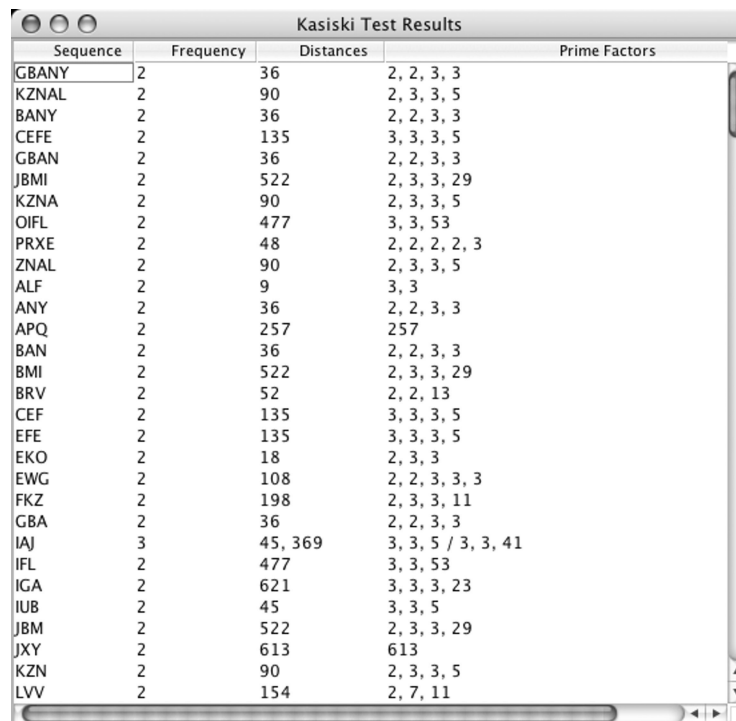


Fig. 2. Data for the monoalphabetic cipher in Figure 1.



Sequence	Frequency	Distances	Prime Factors
GBANY	2	36	2, 2, 3, 3
KZNAL	2	90	2, 3, 3, 5
BANY	2	36	2, 2, 3, 3
CEFE	2	135	3, 3, 3, 5
GBAN	2	36	2, 2, 3, 3
JBMI	2	522	2, 3, 3, 29
KZNA	2	90	2, 3, 3, 5
OIFL	2	477	3, 3, 53
PRXE	2	48	2, 2, 2, 2, 3
ZNAL	2	90	2, 3, 3, 5
ALF	2	9	3, 3
ANY	2	36	2, 2, 3, 3
APQ	2	257	257
BAN	2	36	2, 2, 3, 3
BMI	2	522	2, 3, 3, 29
BRV	2	52	2, 2, 13
CEF	2	135	3, 3, 3, 5
EFE	2	135	3, 3, 3, 5
EKO	2	18	2, 3, 3
EWG	2	108	2, 2, 3, 3, 3
FKZ	2	198	2, 3, 3, 11
GBA	2	36	2, 2, 3, 3
IAJ	3	45, 369	3, 3, 5 / 3, 3, 41
IFL	2	477	3, 3, 53
IGA	2	621	3, 3, 3, 23
IUB	2	45	3, 3, 5
JBMI	2	522	2, 3, 3, 29
JXY	2	613	613
KZN	2	90	2, 3, 3, 5
LVV	2	154	2, 7, 11

Fig. 3. Results of the Kasiski test.

Once the students see how easily they can solve cryptograms ciphered with Vigenère's system on some exercises, the Alberti cipher is introduced along with other techniques, like direct and indirect symmetry [Kahn 1999]. The trial lets the students see how much harder to solve the new cryptograms are and how foolish it was to use the Vigenère square. From this experience, we can conclude Kerckhoffs' most famous maxim *"only a cryptanalyst can judge the security of a crypto system."*

Other polyalphabetic ciphers are discussed in the course and may be used with Ganzúa, for example, the Beauford cipher and the variant on the Alberti cipher attributed to William Friedman, in which an arbitrary alphabet is used for every row in the table. These are simply mentioned in the course, since the former is an instance of the Alberti cipher, and the latter generates cryptograms that are too hard to break in a class exercise.

4. OTHER FEATURES

All the files written by Ganzúa are XML documents and their structures are defined in W3C XML schemata [W3C 2004]. We decided to use the extensible markup language (XML), because it makes data accessible to any user with a text editor and can be easily parsed and validated by other programs. This way, anyone can use the files with the standard relative frequencies of languages or the saved cryptanalysis projects in their own programs without much

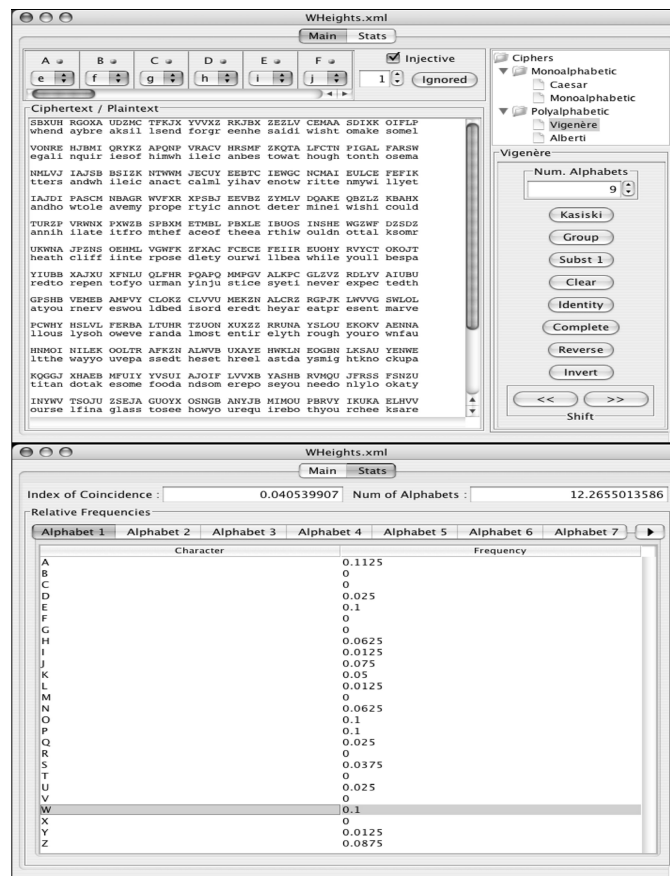


Fig. 4. Tools and data for the Vigenère cipher.

effort. If you teach a course aimed at computer science students, you may want them to use some of those documents in a programming exercise. For example, the files that contain the language frequencies could be used in a program that tries to solve monoalphabetic substitution ciphers with a hill-climbing algorithm.

Since Ganzúa was intended to be a multilingual tool, beyond making it able to handle arbitrary alphabets and many languages, it was also internationalized [Sun Microsystems 2005]. As of version 1.01, it has been localized for English and Spanish, so its interface can be presented on either of these languages, depending on the system's preferences or those of the user. Support for other languages can be added by translating the text files the application gets its labels from. The manual is also available in both languages [Garcia-Pasquel 2004].

The figures presented up to this point show cryptograms in English, with the same plain and cipher alphabet, but Ganzúa can use arbitrary plain and cipher alphabets. Figure 5 shows a cryptogram that uses cyrillic characters in the cipheralphabet and the Spanish alphabet in the plaintext. Using a different

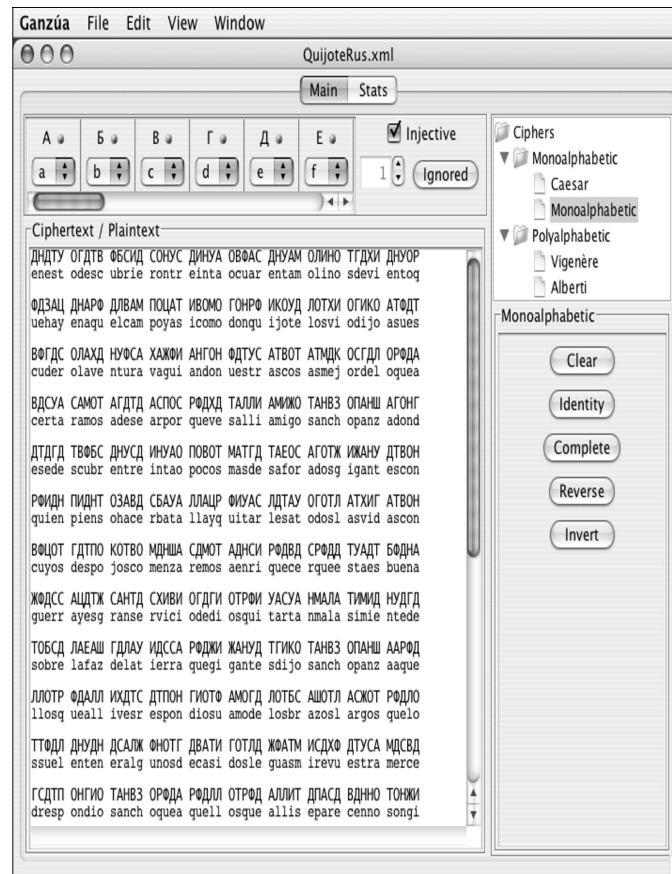


Fig. 5. Spanish text ciphered using cyrillic characters.

character set to cipher is an illusive complication, but can display Ganzúa's flexibility.

5. CONCLUSION

Ganzúa can be a very useful tool during the first stages of an introductory course to cryptology or computer security, particularly to those not in English. The environment, methods, and data it provides, can drastically reduce the time required to solve a monoalphabetic or polyalphabetic cryptogram (approximately 75% in our experience), letting the students concentrate on the cryptanalysis process. Thanks to its use of XML, the data it generates can be used easily in programming exercises.

5.1 Exercises

This section discusses three exercises similar to those mentioned earlier, their objectives, and qualities. We start by describing how to make the examples and continue with the steps that should be taken to find the solution. Before you

try them, we highly recommend installing Ganzúa [Garcia-Pasquel 2004] and reading the manual.

While the cryptograms we use in the course come from text in Spanish and use its alphabet, knowledge of the language can greatly simplify the cryptanalysis. The following examples are in English. The steps described can be applied to other languages, Spanish, in particular.

The main feature of classical cryptosystems is called *confusion*. The cryptanalytic methods used to break them focus on the tracking of elements that cause such confusion. Only the exposure to these techniques may lead to a clear understanding of why most of our modern cyphers also use *diffusion*. The exercises in this section are used to show the students that a cryptanalyst may break a cypher, no matter the amount of confusion used. Once they understand this, the idea of diffusion is introduced.

5.1.1 Monoalphabetic Cipher with Word Separation. This exercise should be used after explaining monoalphabetic ciphers, in general, and as a first exposure to cryptanalysis. It shows how the cryptanalyst exploits data from the original message that remains in the ciphertext and is used to explain a technique that will be used repeatedly to solve monoalphabetic and polyalphabetic ciphers, the frequency analysis.

Write a text file with the message you want the students to find when the cryptogram is solved. The message does not have to be longer than a paragraph. In this example, we will use a fragment taken from Joseph Conrad's *The Heart of Darkness*.

```
No, they did not bury me, though there is a period of time
which I remember mistily, with a shuddering wonder, like a
passage through some inconceivable world that had no hope
in it and no desire. I found myself back in the sepulchral
city resenting the sight of people hurrying through the
streets to filch a little money from each other, to devour
their infamous cookery, to gulp their unwholesome beer, to
dream their insignificant and silly dreams.
Joseph Conrad, The Heart of Darkness.
```

Open the file as ciphertext in Ganzúa and change all the characters to lowercase using the Edit menu. From the same menu, you can remove characters you may not want to appear in the cryptogram, like punctuation marks. You should also add the characters that do not appear in the text, but you would like to consider in the alphabet (q, x, and z in the example). Open the language file with the alphabet you would like to use in the cryptogram (we will use the uppercase English alphabet).

Now we choose the substitution for the characters. Since this exercise is intended to be a first exposure to cryptanalysis and the students already know how the Caesar, Atbash, and mixed alphabet ciphers work, we generally use the Caesar cipher with an arbitrary shift or a combination of Atbash and Caesar, but only tell the students the ciphertext comes from a monoalphabetic substitution. This way, they have to assume mixed alphabets were used, until their analysis tells them otherwise. For this example, we shift the alphabet seven places by

clicking on the Identity button and repeatedly pressing the the left shift tool. The result is the following substitution:

```
no they did not bury me though there is a period of time
UV AOLP KPK UVA IBYF TL AOVBN AOLYL PZ H WLYPVK VM APTL
which i remember mistily with a shuddering wonder like a
DOPJO P YLTLTILY TPZAPSF DPAO H ZOBKKLYPUN DVUKLY SPRL H
passage through some inconceivable world that had no hope
WHZZHNL AOYVBNO ZVTL PUJVUJLPCHISL DVYSK AOHA OHK UV OVWL
in it and no desire i found myself back in the sepulchral
PU PA HUK UV KLZPYL P MVBK TFZLSM IHJR PU AOL ZLWBSJOYHS
city resenting the sight of people hurrying through the
JPAF YLZLUAPUN AOL ZPNOA VM WLWVSL OBYYFPUN AOYVBNO AOL
streets to filch a little money from each other to devour
ZAYLLAZ AV MPSJO H SPAASL TVULF MYVT LHJO VAOLY AV KLCVBY
their infamous cookery to gulp their unwholesome beer to
AOLPY PUMHTVBZ JVVRLYF AV NBSW AOLPY BUDOVSLZVTL ILLY AV
dream their insignificant and silly dreams
KYLHT AOLPY PUZPNUPMPJHUA HUK ZPSSF KYLHTZ
joseph conrad the heart of darkness
QVZLWO JVUYHK AOL OLHYA VM KHYRULZZ
```

The substitution is saved in a new file. This file may be given to the students if they'll be using Ganzúa to assist them, or you could open it the same way you did with the message file and get all the relevant information needed to make this a blackboard example (see Figures 1 and 2).

When the cryptogram is presented to the students, so should the concepts of standard relative frequencies of a language—frequency analysis and contact analysis. For this example, contact analysis will not be as important as the fact that the cryptogram preserves word separation.

```
UV AOLP KPK UVA IBYF TL AOVBN AOLYL PZ H WLYPVK VM APTL
DOPJO P YLTLTILY TPZAPSF DPAO H ZOBKKLYPUN DVUKLY SPRL H
WHZZHNL AOYVBNO ZVTL PUJVUJLPCHISL DVYSK AOHA OHK UV OVWL
PU PA HUK UV KLZPYL P MVBK TFZLSM IHJR PU AOL ZLWBSJOYHS
JPAF YLZLUAPUN AOL ZPNOA VM WLWVSL OBYYFPUN AOYVBNO AOL
ZAYLLAZ AV MPSJO H SPAASL TVULF MYVT LHJO VAOLY AV KLCVBY
AOLPY PUMHTVBZ JVVRLYF AV NBSW AOLPY BUDOVSLZVTL ILLY AV
KYLHT AOLPY PUZPNUPMPJHUA HUK ZPSSF KYLHTZ
QVZLWO JVUYHK AOL OLHYA VM KHYRULZZ
```

First, we focus on the frequency of the characters that appear as one-letter words in the cryptogram (H and P). The most frequent one-letter words in English are a and i. Looking at the standard relative frequencies of characters in English, P looks like a very good candidate for a and H for i. We move on to the two-letter words (UV, PU, TL, etc.), which should have at least one vowel. Since we identified P as a vowel, from PU and UV, we conclude that V should also be a vowel and its frequency suggests that it is letter O. For TL, we look at the frequencies and see that L is the most frequent character in the cryptogram, so it is probably the most frequent character in the English language, e.

Now we look at the three-letter words. The most frequent in the cryptogram is AOL (AOe with our current substitution), which is also the most frequent trigram. In the English language, the most frequent trigram and three-letter word is the.

At this point, it is possible to read some words in the substitution, which makes us think the substitutions have been chosen correctly. Partial substitutions like theYe on the first line or heaYt, on the last one, help find the remaining characters. Some students may notice that the distance between the substitutions chosen so far matches the distance of the characters in the alphabet, thereby solving the cryptogram.

5.1.2 Monoalphabetic Cipher in Blocks. In the previous exercise, one of the most helpful pieces of information left from the original message was the word separation. It made contact analysis almost completely unnecessary. In the following exercise we remove this and present the text in blocks of the same size.

To generate the cryptogram, follow the same steps described for the previous exercise, except perhaps using a mixed-alphabet substitution to complicate it even further. Once you have the substitution, choose to group the ciphertext characters in blocks of an arbitrary length from the Edit menu. You should get something like this:

```
KHQGF DOYOK HQBRM DPFQG HRSQG GFMFY NJLFM YHOHX QYPFA GYWG Y
MFPPF BFMPY NQYZD AYQGG NGRRO FMYKS AHKOF MZYIF JLJNN JSFQG
MHRSG NHPFY KWHKW FYTJB ZFAHM ZOQGG QGJOK HGHLF YKYQJ KOKHO
FNYMF YXHRK OPDNF ZXB JW IYKQG FNFLR ZWGMJ ZWYQD MFNFK QYKSQ
GFNYS GQHXL FHLZF GRMMD YKSQG MHRSG QGFNQ MFFQN QHXYZ WGJZY
QQZFP HKFDX MHPFJ WGHQG FMQHO FTHRM QGFYM YKXJP HRNWH HIFMD
QHSRZ LQGFY MRKAG HZFNH PFBFF MQHOM FJPQG FYMYK NYSKY XYWJK
QJKON YZZDO MFJPN EHNFL GWHKM JOQGF GFJMQ HXOJM IKFNN
```

This exercise, unlike the previous one, will rely heavily on contact analysis, so using the digram and trigram frequencies for the cryptogram and language provided by Ganzúa would be a good idea. Being able to sort them alphabetically and by frequency will prove very useful.

Instead of starting with one-letter words, we try to identify the vowels using relative frequencies of characters and n -grams. The most frequent character is F, while the most frequent trigram is QGF. The frequencies of Q and G are similar to those of t and h, so this is probably the trigram the.

The second most frequent characters in the cryptograms is H, and looking at the digrams, we find fourteen with H to the left and fifteen to the right. Since the contacts are balanced and its frequency is high, H is probably a vowel,² either a, i, or o. The digram ea is very frequent in English, but FH is not and has a frequency similar to that of eo, so we chose o as the substitution for H.

The next character in frequency that has not been assigned a replacement is Y, which also has balanced contacts and could be a or i. The digram ao has a very low frequency in the language, but YH has a considerably higher frequency, similar to that of io, so we choose i.

Continuing in this fashion, and trying to locate possible words in the substitution, we can solve the cryptogram. In this case, it is the same fragment used in the first exercise, but ciphered with a mixed alphabet.

²Consonants typically have a low frequency and unbalanced contacts.

5.1.3 Vigenère Cipher. Solving the previous exercise was a lot more taxing than the first, showing that the less information the cryptogram keeps from the original text, the harder it is to break. This is taken a step further by polyalphabetic ciphers.

The following exercise should be used after explaining the Vigenère cipher and the Kasiski and Friedman Tests. It shows why the techniques used to solve monoalphabetic ciphers cannot be used on their own (or at all) with polyalphabetic cryptograms, but, more importantly, the virtues of the first analytic test, the Index of Coincidence.

To get the cryptogram, write a text file with the message as before. However, this time make it a large paragraph or a couple of paragraphs, since it will be solved by applying frequency analysis to the characters ciphered with each alphabet. If the message is too short, or many alphabets are used, breaking it could become a nightmare. Open the file as ciphertext in Ganzúa, change all the characters to lowercase, remove those that you do not wish to include in the alphabet, and add those that you do, but do not appear in the text. Group the characters in blocks and open the language file with the alphabet to be used in the cryptogram (here we use the uppercase English alphabet).

Choose a short password and set the number of alphabets to its length. Now select the first alphabet, click the Identity button, and shift the substitution until the a is replaced by the first character of the password. Switch to the second alphabet and repeat the process. Do the same for the rest of the alphabets to get a cryptogram like the next one:

```
JCFYH OKICO KSRSS POAAJ VOLFP VVWDK PMVLU KKTGS HFEBV EHWJH
YNHFP IJJWZ WHKSL KSCIW GOWGS PLFCK REFZA RBPHE HMYG UHCJP
OQXNZ YWWHZ ZRHFY SJRUW WHVCW ZWBDL QKWBU JHLWG DZULW ZRDKA
OCNYW ZWHRD PWJPL DLWZH YPPKA ZMPVS XCLEW ZAWIG DJECL DFGJQ
VCQKP VVJZW NSJNU MPWET VWZOE OVLQR ZPGHA FYLSK WZDZV LWGEL
UJKKC JDKWA RYZAP VRXLU NCJNR HAAZR KLOQI FWAJW JPWZA HILQK
ESEEF JAKOF UWOHY LWKSO IXDFZ ALWWA LZPTQ SZFFA RXSOK PUEEH
YTQXE BZEHU KAGWD UABTJ PWJKV YWLKO EOIKJ CMPUL DWJRO GXSRM
RMPHY PLJHW KEOWW TWLLJ OGVCH FAWEE KWEFR DVMNO ENHGB HYPLJ
AAGTU WKJVC PSPHV CLEGG GZVKE PCPWZ WHKSH AJTLD RJEOL YGWNH
YPPAY FFDGF LSUZW ZAGRX HFKCE PJSRS REKGQ UYEWG PVVZO VAFNZ
UDZGF QVHWQ VLVKK IINH KTYFP SJRRY JWNCI EKGQU YERXP VVXRF
HMKZG AOAZD VLDSZ OHSKT CTIWQ DFYWZ AARDL ELCJD LTHSF CLELF
FMDTH SZEK YIITR MOHFC HUWZC DREAC WEKWI SEEDD DOSTW KKTGS
RKARV ADJPS UODQO OKXRK PHVCU WOHIT DDISE QDFYW VOWZA FVXLY
DHSPR LDSIX HFQDF YPSNG GPUZW DJTQX AFZZU LKHYP PKAZM PVSJR
IPDVU HFHHD YCDPD EEGJT RFWFP PQLAF GCLKA MVEDU NCJDW ZAULW
IGBG LFWIW EOVL DOKLUW PCFFU EEBUD DKKII DDJAH FEKGO SFQWZ
APVLV LOHYL WHAFZ DKAJH VWOWY HJGDK POEOF GKZRY GMJGP XSSPV
VELUN SXLUV ARKSL KAOIE KOEHY PQNEC LDHQA GRYGK HCNWB SJRJF
UWHMU CHOPV VTUHH OEDDY WWEDW MO
```

In order to solve the cryptogram, we start by obtaining an estimate of the number of alphabets used based on the Index of Coincidence and perform the Kasiski Test to fine-tune the estimation. Ganzúa shows that the Index of Coincidence of the previous cryptogram is 0.04388, while that of the English language is 0.06436, which yields an estimate of 4.7578 alphabets. After performing the

Kasiski Test, we find that the longest character sequence with repetitions is HYPPKAZMPVS, which appears twice, at a distance of 618 (prime factors 2, 3, and 103). The most frequent sequence is WS, which appears 16 times and many of the distances between the occurrences share the prime factors 2 and 3. This gives us an estimate of six alphabets.

We proceed assuming six alphabets were used, so we set it as the number of alphabets used in Ganzúa. In order to get the characters ciphered with each alphabet organized in columns, we group them in blocks of six. Now we look at the frequency analysis for each of the columns (as shown in Figure 4). Since the Vigenère cipher uses rotated alphabets, we need only find the substitution for one of the characters. Plus, we can use the distribution of the frequencies in the language to identify them. For example z is the least frequent character in the English language and x is very infrequent too, while a is one of the most frequent. By looking for a very frequent character (a) preceded by three with a low frequency (x, y, and z), those at the ends zero or close to it, we can identify the rotation.

In this cryptogram, the relative frequencies of the first alphabet show the W as the second most frequent, while those of V and T are zero, and U is very infrequent. If we set the substitution $W \rightarrow a$, by clicking the identity button and using the shift the alphabet, we can notice that the rest of the characters are replaced by reasonable substitutions as well. For instance, the most frequent character (A) is replaced by e. Proceeding with the rest of the alphabets the same way, we find that W, O, R, L, D, and S were replaced by a and that the text is a fragment from H. G. Wells' *The War of the Worlds*.

```
JCFYHO KICOKS RSSPOA AJVOLF PVVWDK PMVLUK KTKSHF EBVEHW
noonew ouldha vebeli evedin thelas tyears ofthen inetee
JHYNHF PIIJWZ WHKSLK SCIWGO WGSPLF CKREFZ ARBPHF HMRYGU
nthcen turyth atthis worldw asbein gwatch edkeen lyandc
HCJPOQ XMZYWW HZZRHF YSJRUW WHVCWZ WBDLQK WBUJHL WGDZUL
losely byinte lligen cesgre aterth anmans andyet asmort
WZRDKA OCNYWZ WHRDPW JPLDLW ZHYPPK AZMPVS XCLEWZ AWIGDJ
alashi sownth atasme nbusie dthems elvesa boutth eirvar
ECLDFG JQVCQK PVVJZW NSJNUM PWETVW ZOEVL QZRZPGH AFYLSK
iousco ncerns theywe rescru tinise dandst udiedp erhaps
WZDZVL WGELUJ KKCJDK WARYZA PVRXLU NCJNRH AAZRKL OQIFWA
almost asnarr owlyas amanwi thamic roscop emight scruti
JWJPWZ AHILQK ESEEFJ AOKFUW OHYLWK SOIXDF ZALWWA LZPTQS
niseth etrans ientcr eature sthats warman dmulti plyina
ZFFARX SOKPUO EHYTQX EBZEHU KAGWDU ABTJPW JKVYWL KOEOIJ
dropof waterw ithinf initec omplac encyme nwentt oandfr
KCMPUL DWJROG XSRMRM PHYPLJ HWKEOW WTLLJ OGVCHF AWEEKW
ooverth hisglo beabou ttheir little affair sseren einthe
EFRDVM NOENHG BHYPLJ AAGTUW KJVCPS PHVCLL EGGZVK EPCPWZ
irassu ranceo ftheir empire overma tterit isposs ibleth
WHKSHA JTLDRJ EOLYGW NHYPPA YFFDFG LSUZWX AGRXHF KCEPJS
atthei nfasor iaunde rthem croscro pedoth esamen oonaga
RSREKG QUYEWG PVVZOV AFNZUD ZGFQVH WQVLVK KIINH KTYFPS
veatho ughtto theold erworl dsofsp aceass ources ofhuma
JRRYJW NCIEKG QUYERX PVVXRF HMKZGA OAZDVL DSZOHS KTCTIW
ndange rortho ughtof themon lytodi smisst heidea oflife
```

QDFYWZ AARDLE LCJDLT HSFCLF LFFMDT HSZELK YIITRM OHFCHU
 uponth emasin possib leorim probab leitis curiou storec
 WZCDRE ACWEKW ISEEDD DOSTWK KTKSRK ARVADJ PSUODQ OOKXRK
 allsom eofthe mental habits ofthos edepar tedday satmos
 PHVCUW OHITDD ISEQDF YWVOWZ AFVXLY DHSPLR DSIXHF QDFYPS
 tterre strial menfan ciedth eremig htbeot hermen uponma
 NGGPUZ WDJTQX AFZZUL KHYPPK AZMPVS JRIPDV UHFHHD YCDPDE
 rsperh apsinf eriort othems elvesa ndread ytowel comeam
 EGJTRF WFPPQL AFGCLK AMVEDU NCJDWZ AULWIG BGGLFW IWEOVL
 ission aryent erpris eyetac rossth egulfo fspace mindst
 DOKLUW PCFFUE EBUDDK KIIDDJ AHFEKG OSFQWZ APVLVL OHYLWH
 hatare tooorm indsas oursar etotho seofth ebeast sthatp
 AFZDKA JHVWOW YHJGDK POEFGF KZRYGM JGPXSS PVVELU NSXLUV
 erishi ntelle ctsvas tandco olandu nsympa thetic regard
 ARKSLK AOIEKO EHYPPQ ECLDHQ AGRYK HCNWBS JRJFUW HMUCHO
 edthis earthw ithenv iousey esands lowlya ndsure lydrew
 PVVTUH HOEDDY WWEDWM O theirop lansag ainstu s

REFERENCES

- Bauer, F. L. 2000. *Decrypted Secrets, Methods and Maxims of Cryptology*, 2nd. ed., Springer Verlag, New York.
- Bobbitt, M. *Mike's cryptanalysis tool*, <http://cipherlogic.army.ca/cgi-bin/cryptanalysis.pl>
- Cummings, J. *Topics in applied mathematics: cryptography*, Mathematics Department, Carnegie Mellon University, <http://www.math.cmu.edu/users/jcunning/teaching/crypto/>
- Galaviz, J. and Magidin, A. 2003. *Introducción a la criptología*, (course notes in spanish), Departamento de Matemáticas, Facultad de Ciencias, Universidad Nacional Autónoma de México, available at: <http://pateame.fciencias.unam.mx/cripto/index.html>
- García-Pasquel, A. 2004. *Gauzúa: A cryptanalysis tool for classical ciphers*, <http://ganzua.sourceforge.net/en/index.html>
- Kahn, D. 1999. *The Codebreakers*, 2nd. ed., Scribner.
- Kerckhoffs, A. 1883, La cryptographie militaire, *Journal des Sciences Militaires IX*, 5 (Jan.), 161 (Feb.) available in Fabien Petitcolas: <http://www.petitcolas.net/fabien/kerckhoffs/index.html>
- Peschel, J. *Key recovery utilities and resources*, <http://users.castel.nl/~groor01/crack.htm>
- Quer-I, J. 1996. *Criptografía*, Departamento de Computación, Universidad de Buenos Aires, Argentina, course web page: <http://www.dc.uba.ar/eci/96/n1.html>
- Ramió, J. 2003. *Seguridad informática*, Universidad Politécnica de Madrid, Spain, course web page: <http://www.lpsi.eui.upm.es/SInformatica/SInformatica.htm>
- Russell, M. *CRANK: CRyptANalysis toolKit*, <http://crank.sourceforge.net/about.html>
- Sinkov, A. 1990. *Elementary Cryptoanalysis: a Mathematical Approach*, 6th ed., MAA, New Mathematics Library (No. 22).
- Spillman, R. 2002. CAP: A software tool for teaching classical cryptology, *Colloquium for Information Systems Security Education (CISSE 2002)*. Paper available at: <http://cisse.info/CISSE%20J/2002/spill.pdf> further info about the application: <http://www.cs.plu.edu/pub/faculty/spillman/CAP/index.htm>
- Sun Microsystems. 2005, *Core Java internationalization*, <http://java.sun.com/j2se/corejava/intl/index.jsp>
- Weibel, C. 2002, *Math 395: (An introduction to cryptography)*, Mathematics Department, Rutgers University, <http://www.math.rutgers.edu/courses/395/s02/395syllabus.html>
- W3C (World Wide Web Consortium), 2004. *XML Schema: Specifications and development*, <http://www.w3.org/XML/Schema#dev>

Received March 2005; revised April 2006, June 2006, and June 2006; accepted June 2006