BEING COVER AGENT FIXED DELAY, PILOT RIGHT PLANE, CATCH SMALL RADIO

(CODEBUSTERS)

This is the first year CodeBusters will be a National event. A few changes have been made since the North Carolina trial event last year.

1. The Atbash Cipher has been added.
2. The running key cipher has been removed.
3. K2 alphabets have been added in addition to K1 alphabets
4. Hill Cipher decryption has been added with a given decryption matrix.
5. The points scale has been doubled, but the timing bonus has been increased by only 50% in order to further balance the test.

1 TYPES OF PROBLEMS

1.1 ARISTOCRAT (EASY TO HARD DIFFICULTY)

http://www.cryptograms.org/tutorial.php

An Aristocrat is the typical Crypto-quote you see in the newspaper. Word spaces are preserved. No letter will stand for itself and the replacement table is given as a guide (but doesn't need to be filled in by the team to get credit).

FXP PGYAPYF FIKP ME JAKXPT AY FXP GTAYFMJTGF

THE EASIEST TYPE OF CIPHER IS THE ARISTOCRAT

	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
Frequency	4				1	6	3		1	2	2		2			6				3				3	4	
Replacement	I				F	T	A		Y	C	P		O			E				R				H	S	

1.2 ARISTOCRATS WITH SPELLING AND/OR GRAMMAR ERRORS (MEDIUM TO VERY HARD DIFFICULTY)

For these, either words will be misspelled or grammatical errors introduced. From a student perspective, it is what they might expect when someone finger fumbles a text message or has a bad voice transcription. The sentence may contain homophones or odd word breakups. For example, the sentence "Try not to laugh at the sentence or words" may instead be given as:

INL FHZI IAZ QVJWC VI ICX UXHI VH UZNX AZNTU

TRY KNOT TWO LAUGH AT THE SENT AN SORE WORDS

	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
Frequency	2		2			1		3	6	1		1		3			1			1	3	3	1	3		4
Replacement																										

1.3 Patristocrat (Medium to Very Hard Difficulty)

A Patristocrat is just like an Aristocrat, but all the word spaces have been removed. There can be a hint to help get started solving it. Note that occasionally the replacement table characters will encode a word with the remainder of the letters in Alphabetical order. This is known as a K1 alphabet and will be marked as such on the question. A description of the K1 alphabet and a sample solving is at http://toebes.com/Ciphers/Solving%20a%20K1%20Alphabet.htm

EVEFH JXFUL HEFLJ VODHO EXERR FODXV ELDXH DSUYD M

APATR ISTOC RATCI PHERH ASALL THESP ACESR EMOVE D

K1	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
Frequency				5	6	4		4		2		3	1		3			2	1		2	3		4	1	
Replacement	X	Y	Z	E	A	T	F	R	U	I	B	C	D	G	H	J	K	L	M	N	O	P	Q	S	V	W

Instead of encoding the replacement table with a keyword, the original alphabet is encoded with a keyword and the replacement alphabet is Alphabetical order. This is known as a K2 alphabet and will be marked as such on the question. Note very importantly that the frequency is associated with the encoded letter. However, when attempting to reconstruct the alphabet, students need to remember to put the letter that MAPPED to the enciphered letter instead of the reverse. As an aid, when given a K2 alphabet, the frequency box will have the replacement slots at the top.

HKVWV NLAFU TQDQO GASVE DHKVA FEZED QOHQS OV

THEKE YWORD CANAL SOBEI NTHEO RIGIN ALTAB LE

Original	Q	S	T	U	V	X	Z	K	E	Y	W	O	R	D	A	B	C	F	G	H	I	J	L	M	N	P
K2	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
Frequency	3			3	3	2	1	3			2	1		1	3		4		2	1	1	5	1			1

1.4 Xenocrypt (Medium Difficulty)

A Xenocrypt uses a phrase in Spanish. The level of Spanish should correspond to a second-year high school level. All accent marks are removed, but they will encounter words with Ñ. They do not have to put accents on the final answer to count as correct. A K1 or K2 alphabet may be used, and the keyword will be in English.

AIBHJFQ FQNXQAPIPIY LQX FQ FYRISXG

TAMBIÉN ENCONTRARÁS UNO EN ESPAÑOL

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
Frequency	2	1				4	1	1	4	1		1		1			2	5	1	1					3	2	
Replacement																											

1.5 Caesar Cipher (Very Easy)

https://en.wikipedia.org/wiki/Caesar_cipher

In a Caesar Cipher, every character shifts in the alphabet by the same amount. For example, with a shift of 2, A becomes C, B becomes D, C becomes E, etc.

A CAESAR CIPHER SHIFTS EVERYTHING THE SAME AMOUNT.

C ECGUCT EKRJGT UJKHVU GXGTAVJKPI VJG UCOG COQWPV.

1.6 Atbash Cipher (Very Easy)

https://en.wikipedia.org/wiki/Atbash

In an Atbash Cipher, each letter maps to the reverse of the alphabet. For example, with a 26 letter alphabet, A becomes Z, B becomes Y … Y becomes B, and Z becomes A.

THE ATBASH CIPHER HAS A SINGLE MAPPING TABLE.

GSV ZGYZHS XRKSVI SZH Z HRMTOV NZKKRMT GZYOV.

1.7 Affine Cipher (Easy to Medium Difficulty)

http://myothermind.com/affine_cipher.htm

For the Affine Cipher, students may have to encode, decode, or decrypt a phrase. Given any two characters, it is possible to recover the A/B encoding keys. Decoding is best done by encoding known letters. Encoding is simple math. For example, the phrase below is encoded using A=3 and B=7 :

THE AFFINE CIPHER IS A MATH PROBLEM

MCT HWWFUT NFACTG FJ H RHMC AGXKOTR

1.8 Hill Cipher (Medium Difficulty but Tedious)

http://myothermind.com/hill_cipher.htm

The Hill Cipher is a pure matrix math problem using either a 2x2 or 3x3 matrix. Students will only have to encode this, but it must be done in groups of letters (2 or 3 depending on the matrix). The most common mistake is to not pad with the letter Z to make a full matrix for each group. Another common mistake is incorrectly mapping the resulting numbers to letters.

$$\begin{pmatrix} H & I \\ L & L \end{pmatrix} \equiv \begin{pmatrix} 7 & 8 \\ 11 & 11 \end{pmatrix}$$

A	C	I	P	H	E	R	
Q	W	U	T	D	R	H	U

Students may also be asked to decode a Hill Cipher given the decoding matrix. In this case the inverse matrix is:

$$\begin{pmatrix} 25 & 22 \\ 1 & 23 \end{pmatrix}$$

Q	W	U	T	D	R	H	U
A	C	I	P	H	E	R	Z

1.9 Vigenère Cipher (Medium to Hard Difficulty)

http://myothermind.com/vigenere_cipher.htm

For the VIGENÈRE cipher students may have to encode or decode a phrase. They may also have to recover the key given a hint. The most common mistake is skipping spaces for the encoding key.

S	O	L	V	E
L	V	P	Q	M
T	H	E	V	I

S	O	L	V	E
Y	S	Y	Z	V
G	E	N	E	R

S	O	L	V	E
W	Q	T	K	L
E	C	I	P	H

S	O	L	V	E
W	F	H	V	W
E	R	W	A	S

S	O	L	V	E
U	C	Y	N	M
C	O	N	S	I

S	O	L	V	E
V	S	C	Z	H
D	E	R	E	D

S	O	L	V	E
M	B	M	M	I
U	N	B	R	E

S	O	L	V	E
S	Y	L	W	P
A	K	A	B	L

S	O	L	V	E
W	T	Z	M	H
E	F	O	R	D

S	O	L	V	E
W	Q	L	Y	I
E	C	A	D	E

S
K
S

1.10 Baconian Cipher (Easy to Medium Difficulty)

https://en.wikipedia.org/wiki/Bacon%27s_cipher

http://toebes.com/Flynns/Flynns-19250425.htm (if you want some nice historical reading)

The Baconian Cipher has a special 5 symbol alphabet (using A and B) to encode letters. For example, the text CIPHER encodes as

C I P H E R

AAABA ABAAA ABBBA AABBB AABAA BAAAA

However, the encoded text does not need to literally use the latters 'A' and 'B'. the letters used for A must be distinct from the letters used for B. For example lower case letters could be used to denote an A and upper case letters a B.

hovEr cRaft dROVe unDEr laRge Horse

Or one group of letters stand for A and a different set of letters stand for B. For example, if A-M stands for A and N-Z stands for B you can construct a sentence to encode the same phrase.

ELITE COACH CROWD CARRY METAL TABLE

1.11 RSA Cipher (Easy to Medium Difficulty)

https://en.wikipedia.org/wiki/RSA_(cryptosystem)

The RSA cipher is of extraordinary importance to internet communications and is based on number theory (the branch of mathematics that deals with prime numbers and other oddities, such as modular arithmetic). The RSA

cipher is also used extensively in banking and digital signatures. (It is named for its inventors, Rivest, Shamir, and Adleman.)

In real life, typically 300-digit numbers will be used in RSA. That obviously won't work on a Science Olympiad test. Instead, numbers with 2--4 digits will be used, depending on the question. The questions come in three broad categories.

- The first category tests basic understanding of the components of RSA, (``the alphabet soup'') and their proper uses. These are commonly denoted p, q, n, Φ, e, d, m, c. (When using digital signatures, one adds in H, s, t but Science Olympiad won't go that far.) By the way, the choice of letters is consistent across all textbooks, but the capitalization varies considerably.
    - One could ask what each letter is for.
    - One could ask which are public and which are private.
    - One could provide definitions and ask the student to write the correct letter next to each definition.
    - One could provide a situation and ask which letter or letters should be sent.

- The second category tests the algorithmic understanding of what to do, and when, for a given situation. The situation might be simple or complex, and the student has to read, digest, understand, and decide what needs to be calculated. The answer is a formula for what the computer must compute. For example, there might be three boxes, where the student must put the correct numbers into the boxes, in order to complete the formula. This is great for situations where the actual computation would be remarkably tedious without a computer.

- The third category are closer to traditional mathematics problems, where the student is asked to actually produce a final answer. One must understand RSA and modular arithmetic extremely well in order to be able to know what to do, when, and how to do it. This is more about computational tricks (e.g. the “extended Euclidean algorithm,” or “rapid modular exponentiation” which is sometimes called “the method of repeated squaring”) than rote calculation or arithmetic drill. While no one in real life would ever do RSA by hand, true mastery of the underlying knowledge can only occur by ‘opening up the hood’ and seeing how all of this is done ‘behind the scenes’ by computers.

The strongest recommendation for coaches is to examine the suggested resources and textbooks, and to have students study them. The event supervisors have put a great deal of thought into which books to recommend. The sections dealing with RSA are often moderately short and self-contained. It will often be useful to consult old tests.

2 HINTS FOR YOUR TEAM

- Get your calculators early (they are inexpensive) so that the students become comfortable with them. Note that they may NOT use a standard scientific calculator used at other Science Olympiad events.
- Watch the twitter feed @NCSO_cb (https://twitter.com/NCSO_cb)
- Do the Practice Exams.
- Pay attention to question scores to decide what to do.
- Take advantage of the 2-letter mistake rule to speed up. If you are down to two letters on an Aristocrat and you are sure of the answer, move on to the next question.
- The Timed question is critical. Over 1/3 of the teams solved in under 10 minutes and 23 teams saw their ranking improve by 1 or more positions because of the bonus from it.
- Make Practice Samples.
- Use a pencil and paper.
- Learn to guess! Sometimes a quick guess gets you to a result faster. It is ok to backtrack if it doesn't work out.
- Split out the test among students.
- Bring pencils and erasers.
- Practice, Practice, Practice!
- Have Fun!

3 RESOURCES

https://www.sciencenc.com/resources/high-school/codebusters/ – The main NC Science Olympiad site.

Cipher Tools

- https://toebes.com/codebusters/ – has lots of tools for writing exams and solving ciphers.
- http://www.gregorybard.com/cryptogram.html – includes practice problems and suggested textbooks.
- http://www.cryptograms.org/tutorial.php – One of the best tutorials for solving Aristocrats.
- http://www.dcode.fr/tools-list#cryptography – Has a lot of tools for encoding/decoding ciphers.
- https://quipqiup.com/ – Solves any Aristocrat or Patristocrat.

Practice Sample resources

- http://www.cryptogram.org/ – is the website of the American Cryptogram Association (ACA) if you are looking for even more resources or a fun organization to join. Note: I am a member of the ACA and ACA members will be contributing questions for the test and helping run the event.
- http://cryptograms.org/ – Puzzle Baron's site with tons of Aristocrats
- http://www.cryptoclub.org/ – Has sample ciphers to practice on
- https://www.brainyquote.com/quotes/topics.html – Is a great source of quotes to encode. Keep in mind the length of the quotes however.

4 CREATING A TEST

You can use the template from one of the tests on at https://toebes.com/codebusters/ and just replace the questions with your own. An overview of using the tool can be found at https://www.youtube.com/watch?v=pcz_3ql8ebM

4.1 FOR ARISTOCRATS/PATRISTOCRATS

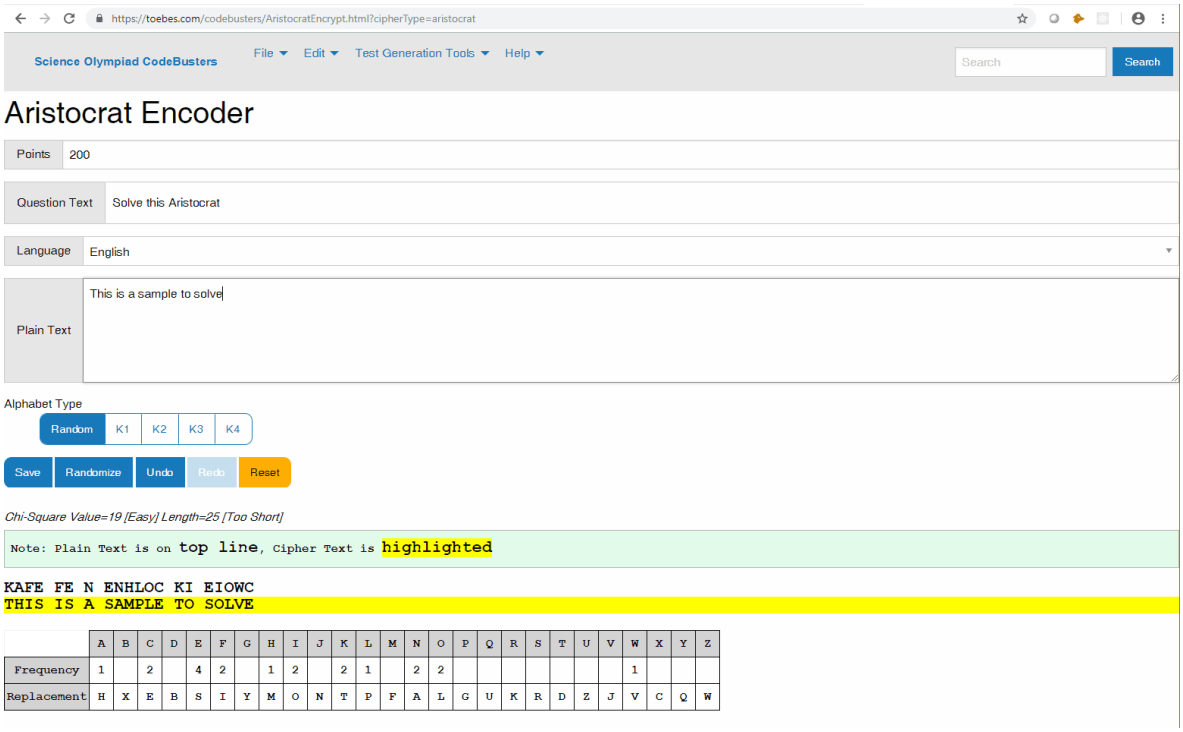
1. Search for Quotes/Phrases to use. Ideally you want something inspirational, topical or science related. A good quote will have around 20 words and about 100-120 characters. They should have a good distribution of letters nominally matching the standard frequency of English letters:

E	T	AO	NIR	SH	LD	CUPF	MWY	BGV	KQXJZ
13%	9%	8%	7%	6%	4%	3%	2%	1%	-

Table 1 - Frequency of English Letters

   The tool automatically checks the phrase and gives a basic idea of difficulty based on a chi-square comparison to the English Frequency. Phrases that start with `it is`, have multiple occurrences of `the` or contain the words `these`, `there`, `little` or `people` tend to be easier. You will also want some samples which have repeated words to use for test questions providing hints. It is good to avoid quotes which are unattributed or by anonymous to allow the author of the quote to serve as an extra hint.

2. Using the Patristocrat or Aristocrat tool https://toebes.com/codebusters/AristocratEncrypt.html (Figure 1) as appropriate, enter the text for the cipher as well as the number of points and the text for the question.



3.

Figure 1 - Aristocrat Encrypt Tool

4.2 For the Spanish Xenocrypt

http://toebes.com/Ciphers/AristocratSpanishEncrypt.html

1. Pick a Spanish phrase which primarily consists of words which a second-year Spanish class would cover. Phrases which have both `la` and `las` present are good choices as well as phrases which contain `y` or Spanish words which are substantially like their English equivalent words are also good. Although it isn't strictly necessary, try to avoid phrases which depend on accented characters. As with the approach for the English Aristocrats, pay attention to the frequency of letters. You can use the Spanish frequency check tool to verify the difficulty.

E	A	O	SNR	IL	DTUC	MP	BHQ	YVGÓÍ	FJZÁÉÑXÚKWÜ
13%	12%	8%	7%	6%	5%	3%	2%	1%	-

2. Encode using the Spanish Aristocrat encoder at https://toebes.com/codebusters/AristocratSpanishEncrypt.html?cipherType=aristocrat. If the encoded string uses both `N` and `Ñ`, you will probably want to re-encode until you don't get them both to avoid confusion on the part of the teams. Although you can also try for an encoding that doesn't use `Ñ` at all, it is perfectly fine to generate a question which has one.

4.3 Hill Cipher

https://toebes.com/codebusters/HillEncrypt.html

1. Pick a phrase to encode. As a rule of thumb for a 2x2 matrix, every pair of letters is worth 20 points. Ideally you want an odd length string to force them to use a padding `Z`. For a 3x3 matrix, every group of three letters is worth 25 points. Again, you want a string which is not a multiple of 3 characters long so that they must add the appropriate number of padding characters.

2. Pick an encoding key. For a 2x2 it is 4 characters long and for a 3x3 it is 9 characters long. This is probably the hardest part to making the test because the matrix must be invertible (https://en.wikipedia.org/wiki/Invertible_matrix). Fortunately, the tool will tell you if it is not invertible. There is also a list of known valid keys at https://toebes.com/codebusters/HillKeys.html for both the 2x2 and 3x3 encodings. In general, it is more likely to be invertible if you use the letters `B`, `D`, `F`, `H`, `L`, `N`, `R`, `T`, `X` and `Z`. as they are odd and non-prime, but you can mix in some other letters. Just make sure that the keyword is not an inappropriate phrase. A total non-sense phrase is perfectly acceptable, but it helps the style of the test if it looks like a word.

3. Use the tool to encode the cipher. The tool can display the math for the problem so that teams can practice and understand what may be wrong with their answers.

4.4 Vigenère Key Cipher Encoding

https://toebes.com/codebusters/VigenereEncrypt.html

1. Pick a phrase to encode. This question is nominally worth one point per letter so a 50-letter phrase (not counting spaces) is ideal.

2. Pick a short 5- or 6-character code word. Ideally you want to have 5 different characters and avoid the letter `A` as it causes a letter to map to itself. If you are doing a Running-Key Cipher, then you can make a phrase as long as you like and not worry about the letter A.

4.5 Vigenère Decoding

1. Pick a code word to use to encode the phrase with. It should be 5 characters long without any repeated letters and doesn't have the letter A in it.

2. Pick a phrase to be decoded. It should be about 50 characters long - the question is nominally worth 2 points per character. It should contain one word that is 7 or 8 characters long that will be identified in the question to the team.

4.6 Misspelleedd[sic] Aristocrat

1. Pick a phrase/quote to encode. Ideally this should contain words which have homophones available. The phrase should be about 120-150 characters long as the question is worth 3 points per letter.

2. Use a homophone generation tool (like http://homophonemachine.allaboutlearningpress.com/ or http://evanshort.name/homophone/) or even try dictating through Siri or Dragon type to get a phrase which has been slightly twisted. You may want to try a couple of times to get something that is appealing. Siri has gotten a lot smarter lately and doesn't make as many mistakes as it used to.

3. Encode like a normal Aristocrat using the Aristocrat tool.

4.7 Affine Cipher Basic Question

https://toebes.com/codebusters/AffineEncrypt.html

1. Pick a 5 or 6 letter word to encode which doesn't have the letter A in it.

2. Pick a value for *a* which is not coprime with 26 (1,3,5,7,9,11,15,17,19,21,23 or 25). The actual value doesn't matter, but larger ones tend to be slightly harder. If you are generating tests for multiple regions, pick numbers that are near each other. I.e. 7, 9 and 11 would be good to have as equivalent *a* values.

3. Pick a value for *b* between 1 and 25 inclusive. Unlike a where the larger values become slightly harder, the value of *b* can truly be any number and be the same level of difficulty.