

Mathematics of Information Processing and the Internet

Author(s): Eric W. Hart

Source: *The Mathematics Teacher*, Vol. 104, No. 2 (SEPTEMBER 2010), pp. 138-143

Published by: National Council of Teachers of Mathematics

Stable URL: <https://www.jstor.org/stable/20876803>

Accessed: 28-09-2019 15:01 UTC

REFERENCES

Linked references are available on JSTOR for this article:

https://www.jstor.org/stable/20876803?seq=1&cid=pdf-reference#references_tab_contents

You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



National Council of Teachers of Mathematics is collaborating with JSTOR to digitize, preserve and extend access to *The Mathematics Teacher*

Eric W. Hart



Mathematics of Information Processing and the Internet



*Access, security, accuracy, and efficiency—
all twenty-first-century high school mathematics
curricula should contain content relating
to these four fundamental themes.*

The Internet is everywhere in contemporary society; it is emblematic of the information age in which we live. We must somehow process all the information that bombards us so that it is manageable and useful. Mathematics can help.

Think about the process of buying a song online. First, you must find the song. Suppose you know that the style of music is electronica and that the artist's name contains the word *mouse* but you cannot remember any more information about

<secure transaction>

<password protection>

<encoded information>



<digital fingerprint>

<virus scan>

<encryption>

the song. You might Google *mouse* to look for the song, but you will get too many results that do not have anything to do with music. It would be more efficient to search for *electronica and mouse*. For many search engines, *and* is the default setting, so you can just search for *electronica mouse*. When you look at the list of results, you see the name of the artist you like—Mouse on Mars.

What does this exercise have to do with mathematics? Internet searches use set theory and logic. For example, the set operation of intersection and the Boolean (or logical) operator *and* are used in this sample search.

Once you find the song, you can buy it and download it. When you buy the song using a credit card, you want to be sure that your credit card number is kept secure. Online security makes use of the area of mathematics called cryptography. After purchasing the song, you want the download to be fast and accurate, and you do not want the file to take up too much space on your iPod. Ensuring an accurate download involves the mathematics of error-detecting and error-correcting codes. To make the download fast and the file size compact requires the mathematics of data compression.

This example illustrates four key themes of information processing, particularly as related to the Internet: access (finding information easily), security (keeping information confidential), accuracy (ensuring accurate information), and efficiency (data compression). Each theme will be briefly discussed with reference to high school mathematics.

ACCESS

To be useful, information must be accessible. Consider searching for information on your computer or on the Internet. A “Quick Tip of the Week” from Apple Hot News on October 21, 2008, stated, “In addition to searching any Mac on your network, Spotlight also lets you take advantage of Boolean search operators—AND, OR, NOT ...” (<http://www.apple.com/business/theater/mac.html#booleansearches>). Similarly, the Library of Congress Web site provides a help page devoted to Boolean searches (<http://catalog.loc.gov/help/boolean.htm>) to help online visitors find information in the library more easily (see **fig. 1**).

Figure 1 could be used as a starting point and real-world application for a lesson on elementary set theory, Venn diagrams, and basic logic. These topics not only are central to information processing and the Internet; they also have applications in many other areas of life and mathematics. They should be part of the high school mathematics curriculum for all students.

SECURITY

For information to be useful, it must be secure. Information security is vital for governments, companies, and individuals. You do not want your credit card number stolen when you make an online purchase. Embassies abroad need to send information back to their home governments securely. An e-mail message sometimes must be confidential. All these goals are achieved through cryptography, the study of mathematical concepts and methods for making information secure. Cryptography is used to design cryptosystems, which work as shown in **figure 2**.

There are two basic types of cryptosystems: symmetric key and public key. In a symmetric-key cryptosystem, the same key is used to encrypt and decrypt. Thus, the security of the system depends on the secrecy of the key. In a public-key cryptosystem, different keys are used for encryption and decryption. One key is made public, and the other is kept secret. Symmetric-key systems are faster whereas public-key systems are more secure, so users often prefer hybrid cryptosystems, in which the same key is used to encrypt and decrypt but the key is shared through a public-key system.

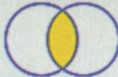


Concept	Search Examples	Retrieval Formula
AND	rodgers AND hammerstein children AND poverty "civil war" AND virginia	 Retrieves only records containing both terms.
OR	sixties OR 60s OR 1960s labor OR labour email OR e-mail OR "electronic mail"	 Retrieves records containing either one or more terms.
NOT	caribbean NOT cuba jockey NOT disc "civil war" NOT american	 Excludes records containing the second term.
NESTING	fruit AND (banana OR apple) (women OR woman) AND basketball ((color OR colour) AND (decorate OR decoration)) NOT (art OR architecture)	Use parentheses () to group portions of boolean queries for more complex searches.

Fig. 1 Boolean searching is well described on the Library of Congress Web site.

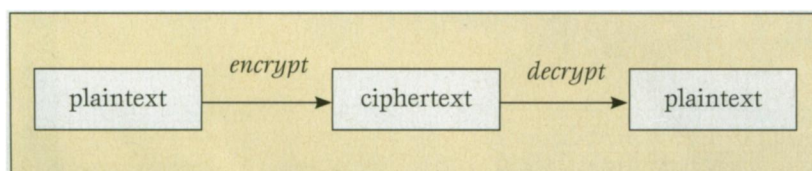


Fig. 2 A cryptosystem uses keys to convert plaintext to ciphertext and then back to the original plaintext.

You often see cryptosystems in action when you use the Internet. For example, when you enter personal information on a Web page, you might see a warning message like that in **figure 3a**. When shopping online, you know that cryptography is being used to keep your transaction secure when you see that the Web site address begins with https instead of http, as in **figure 3b**. This prefix indicates that the Secure Sockets Layer (SSL) protocol is being used to transfer information securely. According to the Apple OS X Help guide, “Web browsers and many websites use the SSL protocol to transfer confidential user information, such as credit card numbers. SSL uses a public and private key encryption system.”

A common example of a symmetric-key cryptosystem is a substitution cipher, in which characters of the plaintext message are replaced by other characters to create the ciphertext. The “Code Crackers” lesson found at NCTM’s Illuminations Web site (illuminations.nctm.org) illustrates a simple substitution cipher. A more secure substitution cipher is a Hill cipher, which uses matrix multiplication to scramble the plaintext message. St. John (1998) provides a thorough explanation of Hill ciphers that is suitable for high school students.

The biggest drawback to symmetric-key cryptosystems is that the common key must be transmitted and kept secure. Before I can send you a secret message, I need to send you the key we will use. But to send you the key securely, I will need to encrypt it with another key, which I will need to send to you. But then we will need a key for that key, and so on.

An elegant solution to this quandary was provided in the mid-1970s with the development of public-key cryptography by Whitfield Diffie and Martin Hellman at Stanford University and Ronald L. Rivest, Adi Shamir, and Leonard Adleman at the Massachusetts Institute of Technology. For example, consider the RSA public-key cryptosystem (the initials in the name honor the MIT developers). This system is based on the facts that multiplying two large numbers—in order to create the encryption—is relatively easy, whereas factoring a large composite number—in order to break the encryption—is difficult. Messages are first converted to numbers (e.g., A becomes 1, B becomes 2, etc.), and then the numbers are encrypted and decrypted using computations done with modular arithmetic, as follows.

To begin, multiply two very large prime numbers, p and q , to generate a large composite number, n . Next, compute $r = (p - 1)(q - 1)$. Then choose a number e (for encrypt) that has a multiplicative inverse, d (for decrypt), under multiplication *modulo* r . Encryption involves raising to the

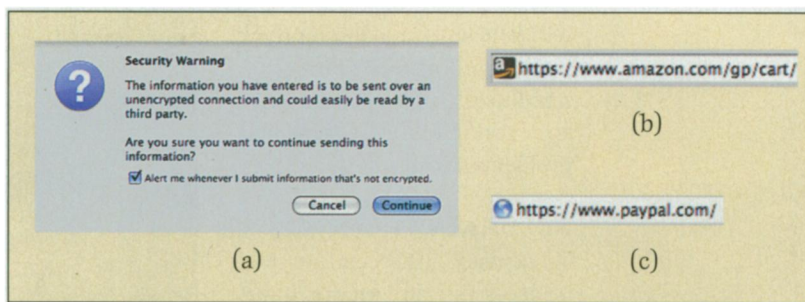


Fig. 3 These examples show the pervasiveness of cryptography on the Internet.

power e and reducing mod n , while decryption involves raising to the power d and reducing mod n . The numbers n and e provide the public encryption key available to anyone, and d is the private key used for decryption.

This procedure works—that is, the decryption undoes the encryption—because of a special case of Euler’s theorem: $(M^e)^d \equiv M \pmod{n}$, where p and q are prime numbers, $n = pq$, $r = (p - 1)(q - 1)$, and e and d are multiplicative inverses mod r .

Suppose Alice wants to send a message to Bob using the RSA cryptosystem. Bob, the receiver, has created public and private keys using the method above. He publishes the public key so that it is available to anyone who wants to send him a message. He keeps the private key secret, known only to himself. Alice uses Bob’s public key to encrypt her message and sends it to Bob. Bob uses his private key to decrypt the message.

Think about the security of this message transmission. Suppose an adversary, Carol, wants to steal the message. Carol already knows n and e , which are publicly known as the public key. To steal the message—that is, decrypt it—Carol needs the private key, which is the number d . To find d , Carol needs to know r , because d is the multiplicative inverse of $e \pmod{r}$. To find r , Carol needs to know p and q , because $r = (p - 1)(q - 1)$. But p and q are the prime factors of n , which Carol already knows. Thus, to break the code, an adversary would need to factor the large composite number $n = pq$. Factoring large numbers is acknowledged to be a hard problem in mathematics, so transmitting information using the RSA public-key cryptosystem is presumed to be secure.

How much of this mathematics is core content for all high school students? We want students to be literate consumers in an Internet-based information age, and we want to keep doors of opportunity open for them as we prepare them for college and the world of work. According to these criteria, it is reasonable to expect all students to learn the basics of modular arithmetic and, in broad strokes, how it is applied to provide information security. Besides the fundamental application to information processing, modular arithmetic is rich mathematics that

will help students deepen their understanding of algebraic structures and properties and strengthen their reasoning and problem-solving skills.

ACCURACY

To be useful, information must be accessible and secure, but it must also be accurate. Error-detecting and -correcting codes help make information accurate. Generally, error-detecting codes are used to help ensure accuracy of identification numbers, such as ZIP codes for mail delivery, UPC codes for consumer products, and ISBN numbers for books. Error-correcting codes are more often used to ensure accuracy when transmitting information, such as satellite or cell-phone transmissions.

Error-detecting codes for identification (ID) numbers typically work by appending a check digit to the ID number. The check digit is chosen so that some modular arithmetic property is satisfied.

For ZIP codes, the check digit is a tenth digit appended to the nine-digit ZIP code so that the sum of all ten digits is equivalent to $0 \bmod 10$ (i.e., a multiple of 10). UPC codes are more complicated. Consider a common twelve-digit UPC code, such as 7-4236521685-5 for a carton of heavy whipping cream. The first eleven digits provide the ID number for the product. The final digit—in this instance, 5—is a check digit chosen according to the following algorithm:

- (i) Add all the digits in odd positions, starting from the left ($7 + 2 + 6 + 2 + 6 + 5 = 28$).
- (ii) Triple the sum from step (i) ($3 \cdot 28 = 84$).
- (iii) Add all the digits in the even positions, starting from the left, including only the eleven digits used for product identification ($4 + 3 + 5 + 1 + 8 = 21$).
- (iv) Add the results from steps (ii) and (iii) ($84 + 21 = 105$).
- (v) Choose the check digit, d , so that $d + (\text{result from step [iv]}) \equiv 0 \bmod 10$. Append the check digit as the twelfth digit ($105 + 5 \equiv 0 \bmod 10$).

Different algorithms and modular systems are used for other ID numbers, such as mod 11 for ten-digit ISBN numbers, mod 9 for some traveler's checks, and mod 7 for many rental car ID numbers.



MODULAR
ARITHMETIC
IS RICH
MATHEMATICS
THAT WILL
HELP STUDENTS
DEEPEN THEIR
UNDERSTANDING
OF ALGEBRAIC
STRUCTURES.

An error in an ID number is detected if the designated computation does not meet the specified criterion. For example, if the computer built into a grocery store checkout scanner carries out the UPC code computation for a particular product and obtains, say, $3 \bmod 10$ instead of $0 \bmod 10$, then an error is detected. When that happens, you hear a beep, and the checker at the grocery store must enter the ID number by hand instead of simply by scanning the product. (Here there is a code within a code; the numbers are encoded as bars that are read by the scanner. Mathematics is used to create bar codes also.)

Codes used for ID numbers typically focus on detecting errors and are not particularly strong for correcting errors.

Codes used for transmitting information, such as satellite transmissions, are designed to both detect and correct errors. You do not want to wait several years for information to arrive from your deep-space probe only to have the computer tell you that an error has occurred; you would like to use a code that permits reasonable error correction. A common type of error-correcting code used in digital data transmission is a linear code. (For high-school-appropriate treatment of linear codes, see Hirsch et al. [2009] and Malkevitch et al. [1991].)

EFFICIENCY

A final issue to consider in information processing is efficiency, which we will consider in the context of data compression. Think about the documents, software, photographs, music, and video that you store on your computer, e-mail to a friend, or download from the Internet. You would like these digital data files to transmit efficiently, download quickly, and not take up too much space on your iPod or hard drive.

All this can be achieved through data compression techniques. You know that data compression has occurred when you see files that end with suffixes such as .jpg (photographs), .mp3 (music), .mpg (video), or .zip (general file compression).

The basic strategy of data compression is to use a variable-length code, in which code words are strings of ones and zeros of different lengths. Variable-length code allows the efficient strategy of encoding more-frequently-occurring data with shorter

code words. Thus, a common character such as e is encoded with a shorter code word (say, 01), whereas an uncommon character such as q is encoded with a longer code word (say, 110101). In fixed-length codes, called block codes, such as the well-known ASCII code used to convert characters to ones and zeros for computer use, each code word has the same length. Through the use of a block code, e and q are each encoded with the same number of bits. Using a variable-length code where e is represented with fewer bits than q compresses data.

One of the earliest variable-length codes, still used as part of most modern proprietary compression algorithms, is a Huffman code. To create a Huffman code, users consider the relative frequencies of characters in the message and construct a specific type of vertex-edge graph, called a Huffman tree, to find the code words. (For a high school lesson on Huffman codes, see Hirsch et al. [2009] or experiment with an Internet applet that illustrates the process of Huffman encoding [e.g., <http://peter.bittner.it/tugraz/huffmancoding.html>].)

We conclude with some recommendations for which topics within the mathematics of information processing deserve a place in the high school mathematics curriculum.

A TWENTY-FIRST-CENTURY HIGH SCHOOL MATHEMATICS CURRICULUM

The mathematics of information processing and the Internet can be organized around four fundamental themes: access, security, accuracy, and efficiency. This mathematics includes many interesting and important topics. Which topics should high school students study?

Three filters can be used to help answer this question: (1) Which topics contribute to the quantitative literacy that all students need in the modern technology-rich, information-dense, “flat world” (Friedman 2008) in which they live? (2) Which topics help keep doors of opportunity open for students as they move into college and the world of work? (3) Which topics are not only fundamental to the mathematics of information processing, especially as related to the Internet, but also valuable mathematical topics in their own right and have other important applications? Applying these three filters yields the following recommendations.

Essential Topics for All Students

- Basic set theory, including the operations of union, intersection, set difference, and set complement, and the subset relation (related to the issue of access in information processing)
- Basic Boolean logic, including AND, OR, and NOT (related to the issue of access in information processing)

- Basic modular arithmetic, including congruence and multiplicative inverses (related to the issues of security and accuracy in information processing)

Valuable Topics for Many Students

- Use of inverse matrices in cryptography, as in Hill ciphers
- Basic ideas of data compression, including variable-length codes versus block codes and the fundamental strategy of encoding more-frequently-occurring data with shorter code words

Interesting Topics for Some Students

- Matrices and linear algebra used to find single-error-correcting linear codes
- Hamming distance, used in linear codes, as an example of a metric in a setting other than traditional geometry
- Number theory used in the technical details of RSA public-key cryptography
- Use of vertex-edge graphs in constructing a Huffman data compression code
- Data compression ideas, such as prefix-free codes and lossless algorithms

Thus, in different ways for different students, the mathematics of information processing and the Internet is essential in a twenty-first-century high school curriculum.

BIBLIOGRAPHY

- “About Secure Sockets Layer.” OS X 10.5.6 Help.
- Friedman, Thomas L. *Hot, Flat, and Crowded: Why We Need a Green Revolution—And How It Can Renew America*. New York: Farrar, Straus and Giroux, 2008.
- Hirsch, Christian, James Fey, Eric Hart, Sabrina Keller, Harold Schoen, and Ann Watkins. *Core-Plus Mathematics, Course 4*. New York: Glencoe/McGraw-Hill, 2009.
- Malkevitch, Joseph, and Gary Froelich. *Loads of Codes*. Bedford, MA: COMAP, 1993.
- Malkevitch, Joseph, Gary Froelich, and Daniel Froelich. *Codes Galore*. Bedford, MA: COMAP, 1991.
- St. John, Dennis. “Exploring Hill Ciphers with Graphing Calculators.” *Mathematics Teacher* 91, no. 3 (March 1998): 240–45.



ERIC W. HART, ehart@mum.edu, is a professor at the American University in Dubai, United Arab Emirates, and a consultant in Fairfield, Iowa. He is interested in high school curriculum and standards, professional development in content and pedagogy, and discrete mathematics.