Lecture 0: People and secrets

Three may keep a secret, if two of them are dead.
Benjamin Franklin, July 1735

0.1 Historical and social context

Secret communication for commercial, military, and diplomatic purposes has a history thousands of years old. A *very* detailed but sometimes tedious reference to this history is [1]. Briefer but perhaps not as careful in places is [2]. Within the last one or two decades huge changes in this "secret writing" have occurred. Partly this is due to the ubiquity of the digital computer and its interconnections via the Internet. But an equal reason is that new ideas and new applications of classical mathematics have revolutionized the field. We will briefly discuss a few of these ideas and their background. Only within the last decade has some of what is presented here found its way into textbooks. Most of these are at the graduate level. Rather recently, several texts for college undergraduates have been written about these topics. This is new stuff, but it is new stuff which is used daily by virtually everyone living in a "wired" society: everyone who has used an ATM and everyone who has made a so-called secure transaction on the Internet. Everyone whose privacy has been endangered by the omnipresence of eavesdroppers should have some familiarity with what is possible. Every person who wants to be technically literate should know something of what will be described below.

0.2 Social and policy questions

Can medical records be private? The reality is that many people who have access to medical files can view the contents. Wouldn't it be terrifically difficult to create technical barriers to prevent this, while still allowing legitimate access? What we will see is that such control is perfectly feasible, and that asserting that it is not possible creates an illusion making debates about such privacy a bit silly.

Should your e-mail be read by the government? Should it be read by computer system operators? Should criminals (kidnappers, terrorists) have the security of exchanging information over public networks with little chance of being spied upon? These questions are real, due to changes in how secret communication can be accomplished. We'll discuss one or two methods of "public key cryptography" which have changed how the world looks at secure communication. This secret writing is very different from what was done during most of the extended history of cryptography: the protocols* described break with tradition and sometimes seem to contradict intuition.

Almost everyone working in cryptography and computer security is aware of the social implications of their work, and has strong opinions. See [3] for sharp discussion about some of the issues in this area. And also see [4] for an analysis of cyberspace and society. The following quote is from [4].

---

* One definition of a *protocol* as used here is "A set of formal rules describing how to transmit data." We'll use it to mean a careful description of computations and procedures.

> Here is something that will sound very extreme but is at most, I think, a slight exaggeration: encryption technologies are the most important technological breakthrough in the last one thousand years. No other technological discovery–from nuclear weapons (I hope) to the Internet–will have a more significant impact on social and political life. Cryptography will change everything.

0.3 How did this all occur?

All of what we'll discuss is comprehensible with little more than mid-level high school algebra. But the way this "algebra" is used is startling and very, very clever. In addition, some key ideas come from the centuries old study of prime numbers. The results used from that area were certainly seen originally as impractical and irrelevant to "real life". Much of what's below could serve as a perfect case study of the use{less|ful}ness of pure math.

0.4 Who invented it and when and how

Fermat (1600's) and Euler (1700's) discovered results in number theory used here. But the applications we'll study were officially invented and, in one still controversial case *patented* (see [5]), only 20 years ago. Or were they? It turns out that much of what was done in public by mathematicians and computer scientists and electrical engineers had probably been anticipated and perhaps used by secret government communications establishments. See [2], chapter 6, for a discussion of some of the secret history.

0.5 Bibliography

[1] David Kahn, *The Codebreakers; The Comprehensive History of Secret Communication from Ancient Times to the Internet*, Scribners, 1996 ($65, but almost 1200 pages long!).

[2] Simon Singh, *The Code Book : The Science of Secrecy from Ancient Egypt to Quantum Cryptography*, Doubleday, 2000 (paperback, $14).

[3] Whitfield Diffie and Susan Landau, *Privacy on the Line*, MIT Press, 1999 (paperback, less than $20).

[4] Lawrence Lessig, *Code and Other Laws of Cyberspace*, Basic Books, 2000 (paperback, $16.00).

[5] The RSA patent is discussed at `http://www.cyberlaw.com/rsa.html`.

Lecture 1: Secret sharing

1.1 A simple exercise

Each student's first and last name can be counted and the result linked together to form an ordered pair. So `Robert Jones` becomes $(6, 5)$ and `Betty Sanderson` becomes $(5, 9)$. Pairs of students should get together and find the equation of a straight line ($y = mx + b$) containing these points. A possible problem: the line is vertical. Solution: split up and find other partners. Find the y-intercept, b, of the line.

1.2 Two people share a secret

We could think of the y-intercept as a "secret" that these two students share. Neither student alone could possibly disclose (by bribery or stringent persuasion!) the shared secret. In fact, if we had a secret b we could take a random m and consider the function $L(x) = mx + b$. We could take a group of people, and issue to each person a distinct x-value and the corresponding value of $L(x)$. Then no one person would be able to give away the secret, and *any two* of these people would easily be able to reconstruct the secret.

1.3 Extension: the Coca-Cola problem

The Board of Directors of the Coca-Cola Company decides to lock up the famous formula for Classic Coke (probably a trademarked name, forgive me lawyers!) in a safe. The combination is a certain number. The board decides that the secret of this number should be shared among the members of the board, but only in a way which needs at least three members to get the secret. Three can be trusted, but *not* two! The board further wants *any* three members to be able to reconstruct the secret. Our job is to describe a protocol for how this can be done.

There are a number of ways to solve this problem. Below is one which generalizes the simple linear technique mentioned above.

Suppose $\mathcal{S}$ is the secret number. Get random numbers A and B, and create the quadratic polynomial $Q(x) = Ax^2 + Bx + \mathcal{S}$. Give each person on the board a random number x and the corresponding $Q(x)$. Then (since, essentially, a quadratic polynomial has three "degrees of freedom", its three coefficients) any three board members working together should be able to find $\mathcal{S}$.

1.4 Contest

There will be a valuable **PRIZE**. I'll give out ordered pairs containing a part of a shared secret, $\mathcal{S}$, shared so that any three people working together can actually reconstruct the secret. Do it, and win the **PRIZE**!

1.5 How many people can share a secret?

Can we generalize this? Given a positive integer, N, is there a protocol which will divide the information about a secret (here, the secret is a number) among a group of people so that *no* collection of $N - 1$ of these people can possibly have enough information to learn the secret, but *any* collection of N of the people will have enough information to learn the secret? Also, the scheme should require a realistic amount of computation.

Following what was done above for $N = 3$, what's needed is a result like the following:

> Given N points in the plane, there is a unique polynomial of degree $N - 1$ whose graph goes through those points. The secret will then be the y-intercept of this graph.

1.6 Can we interpolate a polynomial?

Given the points, finding such a polynomial is an *interpolation* problem. In real life, we are frequently faced with some data points (collected by an experiment, say). If we want to model the data with a function defined by a formula, asking that the formula give the data points *exactly* is called interpolation. Are we sure we can interpolate a polynomial through any finite number of points?

Let's move up a degree and see if we can find a cubic with audience-given data. That is, we'll take 4 points, and find a polynomial $P(x) = Ax^3 + Bx^3 + Cx + D$ which is guaranteed to go through those points.

Suppose the points are $I = (-2, 4)$, $J = (3, 7)$, $K = (9, -5)$ and $L = (12, 14)$. I'll write out a formula for $P(x)$. You may not like it, but it will work.

First we create some auxiliary polynomials.

$$\begin{cases} Q_I(x) = \dfrac{(x-3)(x-9)(x-12)}{((-2)-3)((-2)-9)((-2)-12)} \\ Q_J(x) = \dfrac{(x-(-2))(x-9)(x-12)}{(3-(-2))(3-9)(3-12)} \\ Q_K(x) = \dfrac{(x-(-2))(x-3)(x-12)}{(9-(-2))(9-3)(9-12)} \\ Q_L(x) = \dfrac{(x-(-2))(x-3)(x-9)}{(12-(-2))(12-3)(12-9)} \end{cases}$$

Notice the following which are obtained just by plugging in values and seeing coincidences:

$$\begin{cases} Q_I(-2) = \mathbf{1} & Q_I(3) = 0 & Q_I(9) = 0 & Q_I(12) = 0 \\ Q_J(-2) = 0 & Q_J(3) = \mathbf{1} & Q_J(9) = 0 & Q_J(12) = 0 \\ Q_K(-2) = 0 & Q_K(3) = 0 & Q_K(9) = \mathbf{1} & Q_K(12) = 0 \\ Q_L(-2) = 0 & Q_L(3) = 0 & Q_L(9) = 0 & Q_L(12) = \mathbf{1} \end{cases}$$

So **clearly*** the desired polynomial is just $P(x) = 4Q_I(x) + 7Q_J(x) + (-5)Q_K(x) + 14Q_L(x)$. If you just plug in the values of x, you'll get the y values that were specified. The 0-1 pattern above guarantees this.

This "construction" of a polynomial is called **Lagrange Interpolation**. Please realize that exact polynomial interpolation of many data points representing the results of physical

---

* You must be in a math class when the word "clearly" is used, because no sane person would agree without considerable thought. Oh well, it is nice to know where you are!

experiments is sometimes not appropriate and can lead to strange effects, but here exact representation of information is wanted.

1.7 Just one polynomial?

Suppose P and $\tilde{P}$ are two cubic (third-degree) polynomials that thread through the four points given above. So $P(-2) = 4$ and $P(3) = 7$ and $P(9) = -5$ and $P(12) = 14$ and the same for $\tilde{P}$. How can we compare P and $\tilde{P}$? A simple way is to subtract them. Call the difference polynomial $D(x)$ (D for **d**ifference). We know due to the coinciding values of P and $\tilde{P}$ that D is 0 at -2 and 7 and 9 and 12. Since $D(-2) = 0$, elementary algebra says we can write $D(x) = (x - (-2)) \cdot$ (a quadratic polynomial) (a root allows a polynomial to be factored). But $D(3) = 0$, and the first part $(x - (-2))$ is not 0 when $x = 3$, so the quadratic polynomial must be 0 when $x = 3$. We can divide again, to get $D(x) = (x - (-2)) \cdot (x - 3) \cdot$ (a linear polynomial). And $D(9) = 0$, giving us $D(x) = (x - (-2)) \cdot (x - 3) \cdot (x - 9) \cdot$ (a constant polynomial). Since $D(12) = 0$, plugging in 12 for x shows that the constant is 0! Therefore $D(x)$ is the zero polynomial, and P and $\tilde{P}$ coincide.

1.8 Yes and yes: polynomial interpolation is possible

Given any N points $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$ with $x_1 < x_2 < \ldots < x_N$, there is exactly one polynomial of degree $N - 1$ whose graph contains the points.

If you know linear algebra, then "solving" for the coefficients of the unknown polynomial means finding the solution of a system of linear equations. The coefficient matrix here has a special name: the Vandermonde matrix. Since the determinant of that matrix is not zero (this is stated as a problem in almost every linear algebra text!), there is a unique solution of the system. In this context this result means there is exactly one polynomial interpolating the given values.

1.9 A secret sharing protocol

Suppose you want to "share" a secret among more than N people, so that any N of the people can recover the secret, but no collection of fewer than N people can possibly know enough to recover the secret. Construct a polynomial of degree $N-1$ whose constant term is the secret. Assign each person a distinct number, x, and also give each person the polynomial's value at the assigned x. This information satisfies the secret sharing requirements.

This is called **Shamir's Secret Sharing Scheme**. It and variants of it are widely used – with implementational tweaks, of course. Some of these will be described. A general presentation of this scheme (also called a "threshold" scheme, for the secret is known as soon as a sufficient number of parties have the desired information) is given in chapter 11 of the text [1], usually used for graduate courses or some upper-level undergraduate courses. It might be readable by high school students with some knowledge of linear algebra. Another source is section 18.3 of [2], recommended by the author for advanced undergraduate courses.

1.10 Bibliography

[1] Douglas Stinson, *Cryptography: Theory and Practice*, CRC Press, 1995 (about 450 pages, $80.00).

[2] Paul Garrett, *Making, Breaking Codes: An Introduction to Cryptology*, Prentice Hall, 2000 (about 525 pages, $70.00).

Lecture 2: Modular arithmetic

2.1 What about implementation?

What kind of arithmetic was needed for secret sharing? We used addition, multiplication, subtraction, and division. Almost surely real implementations of secret sharing will use arithmetic done by computer. The arithmetic should be exact, otherwise information might be lost. The arithmetic should be rather fast, otherwise the impatient human beings of the real world won't use it. Also, the numbers involved shouldn't get large because otherwise things might get unwieldy. Here are some candidates for how to do arithmetic. I tested my candidates with the computation of

$$\frac{a^2 + \frac{2}{3}bc}{d^4 - a + b^2}$$

which is certainly rather modest compared to more extended formulas that could occur in real life.

Four-digit floating point If $a = 12.45$, $b = 13.72$, $c = -52.46$, and $d = -9.25$, then the exact value is close to (!) $-.04332\,97773\ldots$. If we use 4 digit floating point computations with rounding after every arithmetic operation, then the fraction's value becomes $-.04563$, rather different. So much information was lost. Also, floating point arithmetic operations take a good deal of time.

Exact rational arithmetic We could try rational arithmetic, done exactly. Suppose $a = \frac{12}{45}$, $b = \frac{13}{72}$, $c = -\frac{52}{46}$, and $d = -\frac{9}{25}$. Then the exact value of the test fraction turns out to be

$$-\frac{487\,52625\,00000}{883\,64148\,87697}$$

which is a version of **expression swell**: that is, the growth of the numbers and other expressions during the computation. More and more storage space is needed, and computations take longer.

I used `Maple` to do these computations. `Maple` has capabilities similar to those of `Mathematica` and `Derive`, with symbolic, numerical, and graphical capabilities. `Matlab` can also be coaxed to do these computations. `Maple`'s advantage is that I have it at home!

2.2 Clock arithmetic*

What time is it? Suppose the answer is 4 o'clock. In 6 hours, it will be 10 o'clock. And 7 hours after that, it will be 5 o'clock (please: temporarily ignore the AM/PM distinction!). Each multiple of 12 is discarded — we only need hourly information from 0 to 11 (here I think of the 12 on the clock face as a 0, please). The sort of arithmetic implied here is called modular arithmetic. This variety is mod 12 arithmetic. So $10 + 7 = 5 \bmod 12$, and $7 \cdot 4 = 4 \bmod 12$, since $28 = 2 \cdot 12 + 4$ and we'll ignore multiples of 12 for mod 12 arithmetic.

---

* Analog clocks, you know, the circular ones with the hands.

We may instead consider the second hand on a clock with each position representing a second labeled, and with the 60 replaced by a 0. Then if we do addition using the motion of the second hand, we would be considering mod 60 arithmetic. In mod 60 arithmetic, by contrast with the situation above, $10 + 7 = 17$. We wouldn't see a difference with conventional arithmetic until bigger numbers are considered. So $45 + 35 = 10$ mod 60, and $3 \cdot 25 = 15$ mod 60.

The book [1] is a long novel romanticizing (!) the last 75 years of the crypto industry, in all its military, diplomatic, and commercial aspects. The chapter *Cycles* (about 25% of the way through the book) has an extensive discussion of modular arithmetic using the spokes of rotating bicycle wheels to motivate subtracting multiples of a fixed number.

2.3 Mod N arithmetic

Here is the proper definition of "mod N arithmetic", where N is a positive integer. Do the arithmetic operations of $+$ and $\times$ on integers, and then divide by N. Discard the quotient and keep only the remainder, an integer between 0 and $N - 1$. This sort of arithmetic is used in cryptography, and more generally in digital signal processing: coding for storage and transmission of information, and transforming information in images (CAT scans, NMR, etc.).

Let us be more precise and look at an example, say with $N = 4$. In class I'll work out the addition and multiplication tables for mod 4 arithmetic, and try to answer questions about the definition. The usual formal properties of $+$ and $\times$ like associativity and commutativity are still valid.

As a larger project, I'll ask the class to complete the addition and multiplication tables for mod 12 and mod 11 arithmetic. We'll see what happens when we try to solve equations.

2.4 Solving linear equations mod N

We will try to solve some linear equations mod 12 and mod 11. Equations of the form $x + 3 = 7$ mod 11 and mod 12 are easy. More interesting equations mod 12 might be

$$7x = 3 \ ; \ 3x = 5 \ ; \ 3x = 3$$

whose solutions are, respectively,

$$\{x = 9\} \ ; \ \text{No solutions (!)} \ ; \ \{x = 1 \text{ or } 5 \text{ or } 9\}$$

and a first reaction may be, "Wow, what's going on!" It certainly is confusing to have simple equations with no solutions or with more than one solution.

On the other hand, the same equations mod 11 have solutions

$$\{x = 2\} \ ; \ \{x = 9\} \ ; \ \{x = 1\}.$$

Each equation has exactly one solution, comfortingly similar to what we're are used to.

By the way, how does one actually "solve" $7x = 3$ mod 11? One way might be to examine the "times 7" row of the mod 11 multiplication table and look for 3. This occurs in the column labeled with 2. Another way might be to find the *multiplicative inverse* of 7 mod 11 – that is, a number $\mathcal{M.I.}$ so that $(\mathcal{M.I.})\cdot 7$ is 1 mod 11. That's 8. Then $7x = 3$ multiplied by 8 becomes $8(7x) = 8(3)$ so $(8\cdot 7)x = 24$ so $1x = 2$, all mod 11 of course. So again 2 is displayed as the solution. We used associativity and multiplicative inverse and identity and all these abstract things. We can also write $8 = \frac{1}{7}$, mod 11, but people rarely do that because it looks very strange.

2.5 Solving linear equations mod P*

How can we relate how the number of solutions behave to properties of, say, the multiplication tables mod 11 and mod 12? If $ax = b$ has two distinct solutions, X_1 and X_2, then there must be some difference between them (say $W = X_2 - X_1$) which isn't 0 and is some integer between 0 and N. Then $aW = aX_2 - aX_1 = b - b = 0$. This means that aW is a multiple of N. So aW must be 0 mod N, and there must be a 0 inside the mod N multiplication table. But that means N can't be prime. Even more is true: if any two entries in the same row are identical, then again we see that N can't be prime. I am trying to make you agree that all of these following properties are equivalent:

- The modulus is a prime number.
- Every linear equation $ax + b = c$ has a solution.
- Every linear equation $ax + b = c$ has exactly one solution.
- Each row and each column of the multiplication table contains all the integers from 0 up to 1 less than the modulus, in some order, each exactly once. All the 0 entries are all along the edge: there are no 0's inside the table.

I have *not* proved that these statements are logically equivalent. I've only tried to make this believable. If you think about the implications for other primes maybe the results are not actually so obvious. For example, 9001 is prime (you can verify this in your head without much effort, actually, if your head is organized correctly). How many numbers are in the mod 9001 multiplication table? Surely 9001^2 numbers – more than a million numbers. If you do agree with the statements written above, then you should recognize that the seven hundred and forty third row of the 9001 multiplication table contains exactly one mention of 236. This is true, both wonderful and weird. If you don't find that result worth pondering, the number 12345678910987654321 is also prime†, and you may consider (!) the statements above for this prime.

2.6 Proof

Let's discuss the word **proof**. Proofs are very important in *Mathematicsland* but sometimes proofs and logical arguments do not impress students. It all seems like a giant house of cards, quite artificial, with little relevance to things of real interest. If this is true for you, I want to disturb your convictions a bit.

---

* Everyone (?) should know that P must refer to a *prime number*. A prime number is a positive integer greater than 1 which can't be written as a product of two integers greater than 1: it can't be "broken" multiplicatively.

† Clearly? No, not at all clearly.

One purpose of these lectures is to convince you that the abstractions of thought here *are* remarkably relevant. The company you have here is good: many mathematicians historically would also have found this unbelievable. Additionally, I sincerely believe that the details cited at the end of the last topic are totally incomprehensible to humans. Perhaps immortal beings can contemplate the multiplication table of 12345678910987654321 and "see" its properties intuitively. To me, intuition and logic acting together, training each other, form an incredibly useful approach to many problems.

2.7 Solving equations: how fast

It turns out to be easy to solve linear equations mod P: $ax + b = c$. This can even be done "by hand" without much effort, even for moderately large P and a, b, c. The key tool is the Euclidean algorithm (yes, a description was given about 2300 years ago!). We don't have time to discuss the Euclidean algorithm, but it really is easy. Symbolic computer programs like `Maple`* use versions of this algorithm, and they can solve equations like $803x = 743$ mod 9001 in essentially no time noticed by the computer (the answer is 8363!) and in about the same amount of time can solve the equation mod 12345678910987654321 (the answer is 10562 24335 22397 49090).

Solving nonlinear equations mod P is *much* harder. Quadratic equations have been extensively investigated, and even they take a great deal of effort. Equations like $2^x = 743$ mod an integer seem to take much more time. In fact, the observed time needed for these equations increases enormously as the number of digits in P grows. There is no known efficient method for solving such equations.

* and `Derive` and `Mathematica` and `Matlab`