# Computer Science 126

General Computer Science  
Fall 2014  
Robert Sedgewick

## Prologue: A Simple Machine

- Brief introduction  
- Secure communication with a one-time pad  
- Linear feedback shift registers  
- Implications

### Who are you?

<table>
<thead>
<tr>
<th>Intended major</th>
<th>Social Sciences</th>
<th>Other Science/Math</th>
<th>Other Engineering</th>
<th>Humanities</th>
<th>CS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Programming experience</td>
<td>none</td>
<td>some</td>
<td>lots</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class</td>
<td>1st year</td>
<td>Sophomore</td>
<td>Junior</td>
<td>Senior</td>
<td></td>
</tr>
</tbody>
</table>

Over 60% of all Princeton students take COS 126
What is this course about?

A broad introduction to computer science.

Goals
- Demystify computer systems.
- Empower you to exploit available technology.
- Build awareness of substantial intellectual underpinnings.

Topics
- Programming in Java.
- Design and architecture of computers.
- Theory of computation.
- Applications in science and engineering.

“Science is everything we understand well enough to explain to a computer.”
— Don Knuth

“Computers are incredibly fast, accurate, and stupid; humans are incredibly slow, inaccurate, and brilliant; together they are powerful beyond imagination.”
— Albert Einstein

The basics

- Lectures. [Sedgewick]
- RS office hours. everyone needs to meet me!
- Precepts. [Gabai, Ginsburg and team]
  - Tips on assignments / worked examples
  - Questions on lecture material.
  - Informal and interactive.
- Friend 016/017 lab. [undergraduate assistants]
  - Help with systems/debugging.
  - No help with course material.

Piazza. [online discussion]
- Best chance of quick response to a question.
- Post to class or private post to staff.

We do not grade on a "curve".

Grades are based on achievement.

Opportunities for us to determine your level of achievement:
- 9 programming assignments.
- 2 written exams (in class, 10/9 and 12/11).
- 2 programming exams (evenings, 10/23 and 12/8).
- Final programming project (due Dean's date – 1).
- Extra credit / staff discretion. Adjust borderline cases.

We do not grade on a "curve".

Course website

http://www.princeton.edu/~cos126

See www.princeton.edu/~cos126 for full current details and office hours.
Programming assignments

are an essential part of the experience in learning CS.

Desiderata

• Address an important scientific or commercial problem.
• Illustrate the importance of a fundamental CS concept.
• You solve the problem from scratch on your own computer!

1. Prologue: A Simple Machine

• Brief introduction
• Secure communication with a one-time pad
• Linear feedback shift registers
• Implications

Things to do before attempting assignment

• Read Sections 1.1 and 1.2 in textbook.
• Read assignment carefully.
• Install Introcs software as per instructions.
• Do a few exercises.
• Lots of help available, don’t be bashful.

http://introcs.cs.princeton.edu/assignments.php
Sending a secret message with a cryptographic key

Alice wants to send a secret message to Bob.
• Sometime in the past, they exchange a cryptographic key.
• Alice uses the key to encrypt the message.
• Bob uses the same key to decrypt the message.

Example 1. Enigma encryption machine [German code, WWII]
• Broken by Turing bombe (one of the first uses of a computer).
• Broken code helped win Battle of Atlantic by providing U-boat locations.

Example 2. One-time pad [details to follow]

Example 3. Linear feedback shift register [later this lecture]

Encrypt/decrypt methods

Goal. Design a method to encrypt and decrypt data.

Example: S E N D M O N E Y

Critical point: Without the key, Eve cannot understand the message.

Q. How does the system work?

A digital world

A bit is a basic unit of information.
• Two possible values (0 or 1).
• Easy to represent in the physical world (on or off).

In modern computing and communications systems, we represent everything as a sequence of bits.
• Text [details to follow in this lecture]
• Numbers
• Sound [details to follow in this course]
• Pictures [details to follow in this course]
• ...
• Programs [profound implications, stay tuned].

Bottom line. If we can send and receive bits, we can send and receive anything.

Encoding text as a sequence of bits

Base64 encoding of character strings
• A simple method for representing text.
• 64 different symbols allowed: A-Z, a-z, 0-9, +, /.
• 6 bits to represent each symbol.
• ASCII and Unicode methods used on your computer are similar.

Example:

01001000 01101001 01100101 01101110 00100000 01101101 01100111 01101100 01100100 00100000 01110010 01100001 01110100 00100000 01100010 01101001 01110010 00100000 01100011 01100100 01110010 00100000 01100011 01101001 01101110 00100000 01100011 01101001 01101110 00100000 01100011 01101001 01101110 00100000 01100011 01101001 01101110 00100000 01100011 01101001 01101110 00100000 01100011 01101001 01101110 01100100 00100000 01101110 01100110 01101001 01110100 01100100 01100101 01100111 01100111 01100010 01101001 01101110 00100000 01101111 01100111 01101001 01101110 01100100 01100101 01100111 01100111 01100010 01101001 01101110 00100000 01101111 01100111 01101001 01101110 01100100 01100101 01100111 01100111 01100010 01101001 01101110 00100000 01101111 01100111 01101001 01101110 01100100 01100101 01100111 01100111 01100010 01101001 01101110 01101111 01101110 01101111 01101001 01101110 00100000 01101111 01100111 01101001 01101110 01100100 01100101 01100111 01100111 01100010 01101001 01101110 01101111 01101110 01101111 01101001 01101110 00100000 01101111 01100111 01101001 01101110 01100100 01100101 01100111 01100111 01100010 01101001 01101110 01101111 01101110 01101111 01101001 01101110 00100000 01101111 01100111 01101001 01101110 01100100 01100101 01100111 01100111 01100010 01101001 01101110 01101111 01101110 01101111 01101001 01101110 00100000 01101111 01100111 01101001 01101110 01100100 01100101 01100111 01100111 01100010 01101001 01101110 01101111 01101110 01101111 01101001 01101110 00100000 01101111 01100111 01101001 01101110 01100100 01100101 01100111 01100111 01100010 01101001 01101110 01101111 01101110 01101111 01101001 01101110

Example: S E N D M O N E Y
One-Time Pads

What is a one-time pad?
• A cryptographic key known only to the sender and receiver.
• Good choice: A random sequence of bits (stay tuned).
• Security depends on each sequence being used only once.

Note: Any sequence of bits can be decoded into a sequence of characters.

Encryption with a one-time pad

Preparation
• Create a "random" sequence of bits (a one-time pad).
• Send one-time pad to intended recipient through a secure channel.

Encryption
• Encode text as a sequence of $N$ bits.
• Use the first $N$ bits of the pad.
• Compute a new sequence of $N$ bits from the message and the pad.
• Decode result to get a sequence of characters.

Result: A ciphertext (encrypted message).

A (very) simple machine for encryption

To compute a cyphertext from a message and a one-time pad
• Encode message and pad in binary.
• Each cyphertext bit is the bitwise exclusive or of corresponding bits in message and pad.

Def. The bitwise exclusive or of two bits is 1 if they differ, 0 if they are the same.
Decryption with a one-time pad

Sending a secret message with a cryptographic key

Alice wants to send a secret message to Bob.
- Sometimes in the past, they exchanged a cryptographic key.
- Alice uses the key to encrypt the message.
- Bob uses the same key to decrypt the message.

Critical point: Without the key, Eve cannot understand the message.

Why does it work?

Crucial property: Decrypted message is the same as the original message.

Let \( m \) be a bit of the message and \( k \) be the corresponding bit of the one-time pad.

To prove: \((m \land k) \land k = m\)

Approach 1: Truth tables

\[
\begin{array}{cccc}
   m & k & m \land k & (m \land k) \land k \\
   0 & 0 & 0 & 0 \\
   0 & 1 & 0 & 0 \\
   1 & 0 & 0 & 0 \\
   1 & 1 & 1 & 1 \\
\end{array}
\]

Approach 2: Boolean algebra

\[
\begin{align*}
(m \land k) &= 0 \\
(m \land 0) &= m \\
(m \land k) \land k &= m \land (k \land k) \\
&= m \land 0 \\
&= m
\end{align*}
\]

A (very) simple machine for encryption and decryption

To compute a message from a cyphertext and a one-time pad
- Use binary encoding of cyphertext and pad.
- Each message bit is the bitwise exclusive or of corresponding bits in cyphertext and pad.

\[
\begin{array}{ccccccccccc}
   x & y & z & w & v & u & t & s & r & q & p \\
   0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
   0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
   0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
   0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
   0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{array}
\]

Decryption with the wrong pad

Eve cannot read a message without knowing the pad.

One-time pad is provably secure (Shannon, 1940s)
- If each pad is used only once,
- AND the pad bits are random,
- THEN Eve cannot distinguish cyphertext from random bits.
Eve’s problem with one-time pads

Eve has a computer. Why not try all possibilities?

**Problem**
- 54 bits, so there are $2^{54}$ possible pad values.
- Suppose Eve could check a million values per second.
- It would still take 570+ years to check all possibilities.

**Much worse problem**
- There are also $2^{4}$ possible messages.
- If Eve were to check all the pads, she’d see all the messages.
- No way to distinguish the real one from any other.

One-time pad is *provably secure*.

Goods and bads of one-time pads

**Goods.**
- Very simple encryption method.
- Decrypt with the same method.
- Provably unbreakable if bits are truly random.
- Widely used in practice.

**Bads.**
- Easily breakable if seed is re-used.
- Truly random bits are very hard to come by.
- Need separate secure channel to distribute key.
- Pad must be as long as the message.

Random bits are not so easy to find

You might look on the internet.

The randomness comes from atmospheric noise... if you trust the internet.

Next: Creating a (long) sequence of "pseudo-random" bits from a (short) key.
A pseudo-random number generator is a deterministic machine that produces a long sequence of pseudo random bits.

Examples

Enigma.
Linear feedback shift register (next).
Blum-Blum-Shub generator.
...

[ an early application of computing ]
[ research still ongoing ]

"Anyone who considers arithmetical methods of producing random digits is, of course, in a state of sin."

— John von Neumann

Deterministic: Given the current state of the machine, we know the next bit.

Random: We never know the next bit.

Pseudo-random: The sequence of bits appears to be random.

Appears to be random??

• A profound and elusive concept.
• For this lecture: "Has enough properties of a random sequence that Eve can't tell the difference."

Which of these sequences appear to be random?

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Random?</th>
</tr>
</thead>
<tbody>
<tr>
<td>10101010101010101010101010101010101010101010101010101010101010101010...</td>
<td>✗</td>
</tr>
<tr>
<td>01010101010101010101010101010101010101010101010101010101010101010101...</td>
<td>✗</td>
</tr>
<tr>
<td>01101100011011001101100110110011011001101100110110011011001101100110...</td>
<td>✗</td>
</tr>
<tr>
<td>100101000101000110001100011000110001100011000110001100011000110001100...</td>
<td>✗</td>
</tr>
<tr>
<td>010010010011110111011110111101111011110111101111011110111101111011110...</td>
<td>✗</td>
</tr>
<tr>
<td>110010010011110111011110111101111011110111101111011110111101111011110...</td>
<td>✗</td>
</tr>
<tr>
<td>0000001011111011111011111011111011111011111011111011111011111011111011111...</td>
<td>✗</td>
</tr>
<tr>
<td>0000001011111011111011111011111011111011111011111011111011111011111011111...</td>
<td>✗</td>
</tr>
<tr>
<td>0000001011111011111011111011111011111011111011111011111011111011111011111...</td>
<td>✗</td>
</tr>
<tr>
<td>0000001011111011111011111011111011111011111011111011111011111011111011111...</td>
<td>✗</td>
</tr>
<tr>
<td>0000001011111011111011111011111011111011111011111011111011111011111011111...</td>
<td>✗</td>
</tr>
</tbody>
</table>

Note: Any one of them could be random!

Linear feedback shift register

Terminology

• Bit: 0 or 1.
• Cell: storage element that holds one bit.
• Register: sequence of cells.
• Seed: initial sequence of bits.
• Feedback: Compute XOR of two bits and put result at right.
• Shift register: when clock ticks, bits propagate one position to left.

An \([11, 9]\) LFSR

0 1 0 1 0 0 0 1 0 1

More terminology

• Tap: Bit positions used for XOR (one must be leftmost).
• \([N, k]\) LFSR: \(N\)-bit register with taps at \(N\) and \(k\).
Linear feedback shift register simulation

<table>
<thead>
<tr>
<th>Time</th>
<th>History of register contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0 1 1 0 1 0 0 0 0 1 0 1</td>
</tr>
<tr>
<td>1</td>
<td>1 1 0 1 0 0 0 0 1 0 1 1</td>
</tr>
<tr>
<td>2</td>
<td>1 0 1 0 0 0 0 1 0 1 1 0</td>
</tr>
<tr>
<td>3</td>
<td>0 1 0 0 0 1 0 1 1 0 0 1</td>
</tr>
<tr>
<td>4</td>
<td>1 0 0 0 1 0 1 1 0 0 0 1</td>
</tr>
<tr>
<td>5</td>
<td>0 0 0 0 1 0 1 1 0 0 0 1</td>
</tr>
</tbody>
</table>

A random bit sequence?

Q. Is this a random sequence?

A. No, it is the output of an [11, 9] LFSR with seed 0110100010!

Self-assessment on LFSRs

Q. Give first 10 steps of [5, 4] LFSR with initial fill 00001.

Encryption/decryption with an LFSR

Preparation
- Alice creates a book of "random" (short) seeds.
- Alice sends the book to Bob through a secure channel.

Encryption/Decryption
- Alice sends Bob a description of which seed to use.
- They use the specified seed to initialize an LFSR and produce N bits.
  [and proceed in the same way as for one-time pads]
Eve's opportunity with LFSR encryption

Eve has computers. Why not try all possible seeds?
• Seeds are short, messages are long.
• All seeds give a tiny fraction of all messages.
• Extremely likely that all but real seed will produce gibberish.

Good news (for Eve): This approach can work.
• Ex: 1-bit register implies 2047 possibilities.
• Extremely likely that only one of those is not gibberish.
• After this course, you could write a program to check whether any of the 2047 messages have words in the dictionary.

Bad news (for Eve): It is easy for Alice and Bob to use a much longer LFSR.

Key properties of LFSRs

Property 1.
• Don’t use all 0s as a seed!
• Fill of all 0s will not otherwise occur.

Property 2. Bitstream must eventually cycle.
• $2^n - 1$ nonzero fills in an $N$-bit register.
• Future output completely determined by current fill.

Key properties of LFSRs

Property 1.
• Don’t use all 0s as a seed!
• Fill of all 0s will not otherwise occur.

Property 2. Bitstream must eventually cycle.
• $2^n - 1$ nonzero fills in an $N$-bit register.
• Future output completely determined by current fill.

Property 3. Cycle length in an $N$-bit register is at most $2^n - 1$.
• Could be smaller; cycle length depends on tap positions.
• Need theory of finite groups to know good tap positions.
Key properties of LFSRs

Property 1.
• Don't use all 0s as a seed!
• Fill of all 0s will not otherwise occur.

Property 2. Bitstream must eventually cycle.
• $2^N - 1$ nonzero fills in an $N$-bit register.
• Future output completely determined by current fill.

Property 3. Cycle length in an $N$-bit register is at most $2^N - 1$.
• Could be smaller; cycle length depends on tap positions.
• Need theory of finite groups to know good tap positions.

Bottom line.
• [63, 62] register generates $2^{63} - 1$ bits before repeating.

Definitely preferable: small cost, huge payoff.

Eve's problem with LFSR encryption

Without the seed, Eve cannot read the message.

Eve has computers. Why not try all possible seeds?
• Seeds are short, messages are long.
• All seeds give a tiny fraction of all messages.
• Extremely likely that all but real seed will produce gibberish.

Bad news (for Eve): There are still way too many possibilities.
• Ex: 63-bit register implies $2^{63} - 1$ possibilities.
• If Eve could check 1 million seeds per second, it would take her 2923 centuries to try them all.

Bad news (for Alice and Bob): LFSR output is not random.

Goods and bads of LFSRs

Goods.
• Very simple encryption method.
• Decrypt with the same method.
• Scalable: 20 cells for 1 million bits; 30 cells for 1 billion bits.
• Widely used in practice. [Example: military cryptosystems.]

Bads.
• Easily breakable if seed is re-used.
• Still need secure key distribution.
• Experts can crack LFSR encryption.

Example.
• CSS encryption widely used for DVDs.
• Widely available DeCSS breaks it!

Prologue: A Simple Machine

• Brief introduction
• Secure communication with a one-time pad
• Linear feedback shift registers
• Implications
LFSRs and general-purpose computers

### Important similarities.
- Both are built from simple components.
- Both scale to handle huge problems.
- Both require careful study to use effectively.

### Critical differences: Operations, input.
- General purpose computer can simulate any abstract machine.
- All general purpose computers have equivalent power (1) [stay tuned].

### Proven questions

Q. What is a random number?

LFSRs do not produce random numbers.
- They are deterministic.
- von Neumann’s “state of sin”: we know that “deterministic” is incompatible with “random”
- It is not obvious how to distinguish the bits LFSRs produce from random,
- BUT experts have figured out how to do so.

Q. Are random processes found in nature?
- Motion of cosmic rays or subatomic particles?
- Mutations in DNA?

Q. Is the natural world a (not-so-simple) deterministic machine??

"God does not play dice."

– Albert Einstein

### A Profound Idea

**Programming.** We can write a Java program to simulate the operation of any abstract machine.
- Basis for theoretical understanding of computation.
- Basis for bootstrapping real machines into existence.

Stay tuned (we cover these sorts of issues in this course).

```java
public class LFSR {
    public static void main(String[] args) {
        int[] a = {0, 0, 1, 0, 0, 0, 1, 1, 0};
        for (int i = 0; i < 2000; i++) {
            a[0] = a[11] ^ a[0];
            System.out.println(a[0]);
            for (int j = 1; j < 11; j++)
                a[j] = a[j-1];
            System.out.println();
        }
    }
}
```

Note: You will write and apply an LFSR simulator in Assignment 5.