Unit 4 - Big Data and Privacy

The data rich world we live in also introduces many complex questions related to public policy, law, ethics and societal impact. In many ways this unit acts as a unit on current events. It is highly likely that there will be something related to big data, privacy and security going on in the news at any point in time. The major goals of the unit are 1) for students to develop a well-rounded and balanced view about data in the world around them and both the positive and negative effects of it and 2) to understand the basics of how and why modern encryption works 3) to prepare students to complete the AP Explore Performance Task.

Chapter 1: The World of Big Data and Encryption

Big Questions

- What opportunities do large data sets provide for solving problems and creating knowledge?
- How is cybersecurity impacting the ever-increasing number of Internet users?
- How does cryptography work?

Enduring Understandings

- 3.2 Computing facilitates exploration and the discovery of connections in information.
- 3.3 There are trade offs when representing information as digital data.
- 4.2 Algorithms can solve many but not all computational problems.
- 6.3 Cybersecurity is an important concern for the Internet and the systems built on it.
- 7.1 Computing enhances communication, interaction, and cognition.
- 7.3 Computing has a global affect -- both beneficial and harmful -- on people and society.
- 7.4 Computing innovations influence and are influenced by the economic, social, and cultural contexts in which they are designed and used.

Week 1

Lesson 1: What is Big Data?

Students are introduced to the concept of “big data,” where it comes from, what makes it “big,” and how people use big data to solve problems, and how much of their lives are “datafied” or could be.

Lesson 2: Rapid Research - Data Innovations

Students "rapidly research" a topic of personal interest and respond to questions about how that innovation produces, uses, or consumes data.

Lesson 3: Identifying People With Data

Students investigate some of the world's biggest data breaches to get a sense for how frequently data breaches happen what kinds of data is lost or stolen.

Week 2

Lesson 4: The Cost of Free

Students examine some of the economic concerns and consumer tradeoffs related to apps and websites that collect and track data about you in exchange for providing you a service free of cost.

Lesson 5: Simple Encryption

Students are introduced to encryption and use a widget to attempt cracking Caesar and random substitution ciphers.

Week 3

Lesson 6: Encryption with Keys and Passwords

Students use a widget to experiment with the Vigenère cipher to learn about the relationship between cryptographic keys and passwords.

Optional Lesson: Hard Problems - Traveling Salesperson Problem

Optional

Students examine a well-known computationally hard problem in computer science, the Traveling Salesperson Problem (TSP). Students solve small instances of the problem, try to formulate algorithms to solve it, and discuss why these algorithms take a long time for computers (and humans) to compute.

Optional Lesson: One-way Functions - The WiFi Hotspot Problem

Optional

Students explore another computationally hard problem - the "Wireless Hotspot Problem" (also know as the vertex cover or dominating sets problem) - to investigate the characteristics of a "one-way function": a problem which is easy to construct in such a way that you know the solution, but is computationally hard to solve.

Lesson 7: Public Key Cryptography

In this big, multi-step lesson, students learn how the basic mechanics and underlying mathematical principles of public key encryption work. Public key encryption is the basis for most secure transactions on the internet.

Lesson 8: Rapid Research - Cybercrime

Research

Students pick a type of cyber attack or cybercrime and do some “rapid research” to learn more about it. The lesson can be used to wrap up the unit or students may optionally complete the Practice PT in the following lesson before moving on to complete the Explore PT.

Week 4

Lesson 9: Practice PT - Big Data and Cybersecurity Dilemmas

Students complete a small research project about a dilemma related to either Big Data or Cybersecurity. The project mimics elements of the Explore Performance Task.

Chapter Commentary

Unit 4 Chapter 1 - What’s the story?

The story of this chapter is about coming to terms with the world of Big Data that we now inhabit, and addressing the new modern dilemmas that come along with it. In many ways, this unit acts as a current events unit, since the daily news is filled with examples: should the government get “backdoor keys” to encryption algorithms in order to unlock a cell phone used by a terrorist? Should a social media site be able to use the data it has about you and your relationships to direct advertising at you, or sell information about you to others? At the end of the unit, students are asked to develop an opinion supported by their own research about a dilemma related to either cybersecurity or personal privacy.

There are two main threads in this unit, which are interwoven: (1) Big Data and (2) Encryption/Security. Since it is nearly impossible to talk about big data without delving into the issues related to security and privacy, it’s a useful time to learn about encryption and how it works. Encryption can be an engrossing subject in its own right, since it involves interesting algorithms, mathematics, and problem solving, not to mention the aspects of societal impact. Indeed, there are entire undergraduate degrees on the subject - The main goal of our encryption lessons is to work up to understanding how public key encryption works, including the primary mathematical principles that make it possible for two people (or computers) to send encrypted messages to each other over the internet in a way that only the intended recipient can decrypt it.

Ready for the Explore PT?

We think that the end of this unit represents a minimum point at which students could complete a successful Explore performance task. Check out the Performance Task pacing section on page 32 for more details.

Our Approach to the Content

Many of the lessons in this unit are designed as**practice for elements of the Explore Performance Task** . In particular, lesson 2 “Rapid Research” is good practice for the relatively quick research and writing students will have to do for the Explore PT. The goal is for students to become adept with looking up sources, reading/skimming articles for their main points, and being able to explain both sides of an argument or dilemma related to big data, security and privacy. Since issues about personal privacy and security affect students’ daily lives, the research should be **relevant and engaging** for students.

Most of the activities in these first two weeks call for students to be engaged in online research, to use online tools to investigate issues, as well as to discuss and write about the issues. These lessons in particular are a great entry point for the teacher to assume the role of **lead learner**. The first week in the unit will feel like: **big data is great!** The second week, however, might feel like: **big data is scary!** Especially in the latter case, the teacher might need to

attend to their students' runaway paranoia about the harmful effects of big data. We want to present a balanced view, but it is a dilemma! Big data is both great and scary at times. Knowledge and awareness are the best tools to protect yourself.

The activities in the third week around data encryption should look and feel similar to lessons from Units 1 and 2. The general pattern is to introduce a concept through an unplugged activity or thinking prompt, and then "plug it in" by using a **widget** to explore the concept further. The purpose of the widgets is to allow students time to play with some of the encryption ideas, which are often mathematical in nature, to get sense for how they work. We want to give students space to be curious and use the tools to **experiment, explore and discover** some essential properties of encryption.

We encourage teachers to use the practice PT to dry-run some of the procedures and processes for the Explore Performance Task. In particular: put time constraints on the research and writing, grant students latitude to research what they want, monitor the appropriateness of students' research choices, and help them get "unstuck," or push them to be more specific, by appealing to the scoring guidelines and rubric.

Lesson 1: What is Big Data?

Overview

In this lesson, students are introduced to the concept of "big data," where it comes from, what makes it "big," and how people use big data to solve problems. Students are asked to consider how much of their lives are "datafied" or could be, and the teacher will show the projected growth of data in the world. Students will then investigate a big data tool in pairs to evaluate the tool for its usefulness and investigate the source of the data used to make the tool. A key take-away from the lesson is that different considerations need to be made when trying to look at, use, or analyze tools that use big data. The world of big data is big, and we've only begun to figure out how to solve problems with it.

The lesson concludes with a brief introduction to the AP Explore Performance Task which students are recommended to complete at the end of the unit.

Purpose

Big data is a big deal right now, both in the field of computer science and more broadly across fields and industries. Understanding the types of things that can be captured in data and anticipating the types of innovations or new knowledge that can be built upon this data is increasingly the role of the computer scientist. A first step toward understanding big data is a survey of how big data is already being used to learn and solve problems across numerous disciplines. The scale of big data makes it hard to "see" sometimes, and techniques for looking at, working with, and understanding data change once the data is "big." Everything, from how it's stored to how it's processed to how it's visualized, is a little different once you enter the realm of big data.

Agenda

Getting Started (20 mins)

- **Video: Big data is better data**

Activity (30 mins)

- **Exponential Growth and Moore's Law (10 mins)**
- **Big Data Sleuth Card (20 mins)**

Wrap-up (20 mins)

- **Big Data Wrap Up (10 mins)**
- **Introduce Explore PT (10 mins)**

Extended Learning

Assessment

Objectives

Students will be able to:

- Identify sources of data produced, used, and consumed by a web application.
- Given a tool that provides access to a large dataset, explain the kinds of problems such a tool could solve.
- Use a tool that provides access to "big data" and investigate its sources.
- Explain that new techniques are necessary to store, manage, transmit, and process data at the scale it is currently being produced.

Links

Heads Up! Please make a copy of any documents you plan to share with students.

For the Teacher

- **Big data is better data - TED talk** - Video

For the Students

- **College Board - Assessment Overview and Performance Task Directions for Students**
- **Activity Guide - Big Data Sleuth Card** - Activity Guide **Make a Copy**
- **Unit 4 on Code Studio**

Vocabulary

- **Big Data** - a broad term for datasets so large or complex that traditional data processing applications are inadequate.
- **Moore's Law** - a predication made by Gordon Moore in 1965 that computing power will double every 1.5-2 years, it has remained more or less true ever since.

Teaching Guide

Getting Started (20 mins)

Video: Big data is better data

- **Video: Big data is better data - TED talk - Video**

Prompt: Based on what you saw in the video, what is big data?

Discuss: In small groups, have students share their responses. Afterwards, open the discussion to the whole class. The main points to draw out from this conversation are:

Big data means different things, at different times, to different people.

- It can mean devices that are constantly collecting data.
- It can mean digitizing data that's been around for a long time (e.g., every book ever written).
- It can mean machine learning and artificial intelligence.

Discussion

Goal: Get students acquainted with the world of big data. Do some simple investigations into some tools that use big data to get a sense of where the data comes from and how it's used.

Teaching Tip

Timing: This is a rather long Getting Started activity, due to the length of the video. Note that the main activity is shorter to compensate.

Activity (30 mins)

Exponential Growth and Moore's Law (10 mins)

Display: Direct students to the graphic showing the exponential growth of data, either by projecting it or having them find it on the Code.org website.

Remarks

> Part of what contributes to data being "big" is the sheer growth of the amount of data in the world. Let's have a look at a graph that shows us just how large big data is.
>
> As you can see from the chart, the amount of data flying around is growing exponentially, doubling every two years or so. Here's a way to think about how fast this is: The world will produce as much digital data over the next 2 years, as currently existed in all of humanity prior to that. And it will do the same the 2 years after that. And so on. That's a lot!

Moore's Law

Briefly introduce Moore's Law as simple piece of vocabulary.

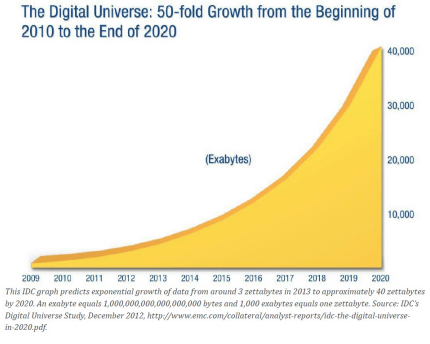
However you do it, here are some key ideas students need to know about Moore's Law:

- Moore's law is actually about computing power, not data, but data growth seems to following the same trend

Teaching Tip

Time check: You should spend at most 10 minutes with exponential growth and Moore's law. Most of the time should be spent working on the "Big Data Sleuth" activity that follows.

Data Graphic: (Note: This graphic is available for students on Code.org website) The IDC's "Digital Universe" is described as "a measure of all the digital data created, replicated, and consumed in a single year." source: **http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf**



This IDC graph predicts exponential growth of data from around 3 zettabytes in 2013 to approximately 40 zettabytes by 2020. An exabyte equals 1,000,000,000,000,000,000 bytes and 1,000 exabytes equals one zettabyte. Source: IDC's Digital Universe Study, December 2012, http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf.

- So far, computing power/capacity seems to double every 1.5-2 years...
- That means it grows exponentially...
- Expontential growth is hard for humans to fathom...
- Yet we need to plan for it.

Moore's Law Remarks

There is a principle in computer science known as **Moore's Law**.

Wikipedia: Moore's Law

It is not a law of nature or mathematics but simply a surprisingly accurate prediction that was made a long time ago. In 1965, a computer chip designer named Gordon Moore predicted that the number of transistors one could fit on a chip would double every 18 months or so.

Amazingly, that prediction has more or less held true to the present day! The result is that since about 1970, computers have gotten twice as fast, at half the cost, roughly every 1.5-2 years. With some small differences, the same is true for data storage capacity.

This is extraordinarly fast growth - we call it exponential growth. With more and more machines that are faster and faster, the amount of data being pushed around, saved, and processed is growing exponentially. This is so fast that it's hard to fathom and even harder to plan for. For example:

- If the average hard drive today is 1 TB and you are planning for something or 6 years away, you should expect that average hard drives will be 8-10 TB.

Key Takeaway: We need to keep Moore's Law in mind as we plan for the future.

Teaching Tip

Keep It Simple: Students only need to have a sense of what Moore's Law is for purposes of putting a name to a description of how fast computing capacity grows, and to understand what it means if they come across it when reading something.

The only reference to it in the framework is this:

`7.2.1F Moore's law has encouraged industries that use computers to effectively plan future research and development based on anticipated increases in computing power.`

Teaching Options:

- Presentation: You can simply give the remarks provided
- Rapid Reseach: Have students do some "rapid research" on Moore's Law and come back with some answers to collectively discuss.

Content Corner

Moore's Law: While the exponential growth principle is the same for data storage as it is for computing power/speed, it's worth noting that Moore's Law actually never referred to data storage capacity, only the number of transistors on a chip.

However, the phrase "Moore's Law" has come to be used colloquially to refer to the idea that in computers and information technology everything seems to double - speed, size, capacity - every 1.5-2 years.

Big Data Sleuth Card (20 mins)

Distribute: Activity Guide - Big Data Sleuth Card - Activity Guide .

Remarks

Big data surrounds us but it is sometimes surprisingly challenging to get access to it, use it, or see it. Much of the data out there is in the "wild." Even when the data is "available," it can sometimes be challenging to figure out where it came from, or how to use it.

Put students into pairs and assign each pair one of the 5 websites listed.

1. Web archive http://www.archive.org
2. Measure of America http://www.measureofamerica.org/maps/
3. Wind Sensor network http://earth.nullschool.net/
4. Twitter sentiment https://www.csc.ncsu.edu/faculty/healey/tweet_viz/tweet_app/
5. Alternative Fuel Locator http://www.afdc.energy.gov/locator/stations/

Student tasks are to follow the resource and answer prompts related to:

1. the visualization tool provided
2. the original source of the data
3. evaluating the usefulness of both the data and the visualization.

Wrap-up (20 mins)

Big Data Wrap Up (10 mins)

Share: Students should share their results from the Big Data Sleuth Cards with members of another group. This can also be conducted as a classwide discussion.

- What kinds of data are out there?
- What format does it come it?
- Where does it come from?
- Did anyone find a link to an actual data source?
- Did anyone find an API? What's an API?

Prompt: After your explorations what do you think "big data" actually means? What makes it "big" as opposed to not?*"

Discuss: Have students share their thoughts with a neighbor. Then share more broadly with the class.

Discussion

Goal: Aim to develop some fluency in talking about the different kinds of data that are available and how they are being used. Students also have an opportunity to assess the usefulness of data visualizations in a new context, now that the scale of data will be much larger.

This is a tricky question that a) doesn't have a fixed definition - whether data is "big" often depends on the context of the data itself or how it's trying to be used and b) even experts might have difficulty pinning it down.

Try to coax student responses toward ideas that discuss using different, new, or unheard of methods or techniques for extracting information from data. The amount of data you have can lend itself to new techniques and this is often what people mean when talking about the "awesomeness" of Big Data.

Remarks

> Here is a general-purpose definition of Big Data (taken from **Wikipedia: Big Data**): "Big data is a broad term for datasets so large or complex that traditional data processing applications are inadequate." The fact that big data is increasingly important across industries reflects rapid changes in how much data we're collecting, and the ways we're using it.
>
> In this unit we're going to be looking into how growth in data and computing more generally is impacting society. In almost every industry and every aspect our lives, computing and data is affecting our lives in both positive and negative ways. This will also be very useful preparation as we begin to look towards the Explore PT.

Introduce Explore PT (10 mins)

Distribute: Give students digital or printed copies of **College Board - Assessment Overview and Performance Task Directions for Students**. We will review pages 4-6 which introduces the Explore PT Components (Digital copy linked to from student resource section for this lesson on the Code.org website).

Teaching Tip

You should be aware that there is a full lesson devoted to Explore PT prep available in the "AP Explore PT Prep" unit. That lesson:

- Does a deeper dive into the task
- Looks at the scoring guidelines
- Reviews scored samples of projects
- Begins the process of preparing to do the real task.

You might want to find a time between this lesson and the end of Unit 4 to take a look at it with your students so they can have some time for the task to sink in.

Remarks

> At the end of this unit we will be doing the Explore PT. To practice the different components of the PT we'll be practicing them throughout this unit. We're going to quickly review those components now, but we'll have opportunities to review and practice them in the next few lessons as well. For right now you don't need to understand all the details, just the big picture.

Review: Quick skim this document with the class, touching on the following points.

- Page 4: The Explore PT has 2 major components, 1. computational artifact, 2. written responses
- Pages 5-6: Skim the submission requirements and give students time to read prompts 2a - 2e.
- Highlight prompts 2c and 2d which references beneficial / harmful effects and the way computing innovations use data, themes of this unit.

Discuss: Respond to any questions students share. Don't lose too much time here. You'll have many opportunities to review the Explore PT in later lessons.

Extended Learning

Open Data: You might be interested in looking at some of the publicly available datasets provided at these sites. It can take a little digging, but you can see the raw datasets and some of the applications that have been made from them.

- **Data.gov**
- **Open Data (opendata.socrata.com)**
- **Open Data Network (www.opendatanetwork.com/)**

Google Maps Traffic: Another big data resource that students may use every day:

- Go to **maps.google.com** and zoom in on your town or city.
- Turn on the Live Traffic view for your area or a nearby town or city.
- The map should show real-time traffic data.
- Have students respond to the same set of questions that they did on the Big Data Sleuth Card. This may take a little more research, since the sources of the data aren't as clearly marked.

Explore PT discussion

Goals:

- Students are aware they will be completing the Explore PT after the conclusion of the Unit
- Students are aware that they will need to 1) research an innovation of interest 2) create a computational artifact 3) respond to written reflections
- Students are aware they will practice these skills throughout and at the conclusion of this unit.

Keep it short: Avoid making this a 45-minute deep-dive review of the AP Explore PT and all its components - a surface level understanding is all that is needed to proceed. Practice questions and direct references to College Board materials are included in each of the next few lessons. This is a kickoff discussion, and you will have many opportunities to prepare students for the PT in later lessons.

Assessment

TBD

Standards Alignment

Computer Science Principles

- **3.2** - Computing facilitates exploration and the discovery of connections in information.
- **7.2** - Computing enables innovation in nearly every field.
- **7.5** - An investigative process is aided by effective organization and selection of resources. Appropriate technologies and tools facilitate the accessing of information and enable the ability to evaluate the credibility of sources.

Lesson 2: Rapid Research - Data Innovations

Overview

In this lesson students will conduct a small amount of research to explore a computing innovation that leverages the use of data. Students will research a topic of personal interest and respond to questions about about how that innovation produces, uses, or consumes data. The lesson is intended to give students practice with doing research of this nature and provides a small amount of scaffolding to help students figure out what to look for.

This lesson is intended to be a quick, short version of a performance task in which students rapidly do some research and respond in writing. It might take 2 class days but should not take more. The goal is to generate ideas for exploration later when students complete the actual Explore PT later in the year.

Purpose

Being able to research modern computing innovations and gain insight into how those innovations are using data is a key skill of computer scientists. This is the first lesson in which students are asked to look at how data is used in a modern computing innovation. Students will learn to look at how data is used with an increasingly critical eye, but this lesson merely sets the table. Having intuitions about how data is used, or how it's not used, can improve one's judgment about modern technology and other innovations that increasingly use, produce, and rely on massive amounts of data to do their work.

Agenda

Getting Started (10-15 mins)

Video - Motivating the research

Activity (30 + 40 mins)

Rapid Reseach - Data Innovations
Day 1 - Choose Innovation, Read and Research
Day 2 - Prepare one-pager

Wrap Up

Presentation (Optional):

Assessment

Objectives

Students will be able to:

- Identify a suitable computing innovation for a research project.
- Identify reliable and authoritative sources of information about a computing innovation.
- Synthesize information taken from multiple online sources to create a cohesive description of a computing innovation.
- Explain how data drives a specific innovation, both in writing and visually.

Links

Heads Up! Please make a copy of any documents you plan to share with students.

For the Students

- **Data Innovation One-Pager** - Template **Make a Copy**
- **Rapid Research - Data Innovations** - Activity Guide **Make a Copy**
- **Computer Science is Changing Everything** - Video (**download**)
- **Data and Medicine** - Video (**download**)
- **The Math Behind Basketball's Wildest Moves** - Video (**download**)
- **[Deprecated] Activity Guide - Rapid Research - Data Innovations** - Activity Guide **Make a Copy**
- **[Deprecated] Worksheet - Data Innovation One-Pager Template** **Make a Copy**
- **Unit 4 on Code Studio**

Vocabulary

- **One-pager** - A business/corporate term for a one-page document that summarizes a large issue, topic or plan.

Teaching Guide

Getting Started (10-15 mins)

Video - Motivating the research

Goal: Develop some ideas (and excitement) about the rapid research project

Video: Show, or have students watch, **one** of the following videos:

- **Data and Medicine - Video** (6:07)
- **Computer Science is Changing Everything - Video** (4:33 NOTE: this video was also shown in Unit 1)
- **The Math Behind Basketball's Wildest Moves - Video** (12 mins)

The purpose is simply to motivate the upcoming research.

Remarks

> One of the things that many modern innovations have in common is their use of data (often Big Data, but not always). To explore how innovations use data more in depth you will be completing a **rapid research project** on a "data innovation" of your choosing.

Get excited! This is your opportunity to dig deeper into a computing topic that has piqued your interest over the entire course.

- What kinds of things are you interested in?
- How does computing affect them?
- How is data used to make innovations you're interested in actually work?

The project mimics some of the things you have to do for the **Explore Performance Task** and will be useful preparation. In particular the Explore Performance Task asks you to:

- Research a modern computing innovation.
- Explain how it uses, produces, or consumes data.

This is exactly what you'll be doing today!

Activity (30 + 40 mins)

Rapid Reseach - Data Innovations

Distribute: Rapid Research - Data Innovations - Activity Guide and **Data Innovation One-Pager - Template** and review as a class.

Below is a suggested schedule for completing the project.

Day 1 - Choose Innovation, Read and Research

Review Activity Guide and Rubric:

At the beginning of the project, emphasize the importance of reviewing the **one-pager template** and **rubric**. Students may assume that more is required of them than is actually the case.

Teaching Tip

Differences from the actual Explore PT: The actual Explore Performance Task will be completed over 8 class hours. The fact that this schedule is significantly shorter reflects several differences in this Practice PT.

- Some categories and topics have been supplied ahead of time.
- The visual students are using does not have to be an original computational artifact

In particular, emphasize that they do not need to create their artifact themselves, but it must still meet the requirements of this project. Point out that the written component is quite short. They probably have space for 100-150 words per response.

Choosing Your Innovation: It is recommended that you place a time limit on this process (e.g. 20 minutes). Students should not leave class after the first day without a topic in mind and ideally with some resources identified. Luckily, in choosing their topics, students will likely have begun to identify resources they can use in completing their project.

Conducting Your Research: This document is intended to serve primarily as a guide to students for identifying online sources of information. The skill students need to develop is identifying useful resources on their own and then synthesizing this information. Being presented with a structured way of doing this means students will have a model for how to complete their research when completing the actual Explore PT.

The "Key Information to Find" highlights specific terminology from the Explore PT that students will benefit from having seen earlier in the course.

Day 2 - Prepare one-pager

Identify a Visual: Students need to identify a visual artifact (image, visualization, drawing, chart, video, interview, etc.) that gives some additional insight into their innovation. Students DO NOT need to make this visual themselves. The goal is to effectively use a visual to communicate information about a technical topic.

Complete One-Pager: Students should find this aspect of their project most familiar. The prompts are similar in style and content to prompts students have already seen. Emphasize the need for clarity in their writing, and remind them that everything must fit on a single page. If they have responded completely to each of the prompts, it is fine to write less.

Sharing/Submission: You may want to collect students' one-pagers, have them share in small groups, or with the whole class. Since students were researching something of their own choosing, they might be eager to show what they found out.

Wrap Up

Presentation (Optional):

If time allows, students may wish to have an opportunity to share their one-pagers with one another. Consider other options like creating a "Data Innovations Museum" by posting links to all their documents in single shared document. Or even print them out and post them in the room to be reviewed in a gallery walk.

Assessment

- Use the rubric provided with the Activity Guide to assess the one-pagers.

Standards Alignment

Computer Science Principles

▸ **1.2** - Computing enables people to use creative development processes to create computational artifacts for creative expression or to solve a problem.

▸ **3.2** - Computing facilitates exploration and the discovery of connections in information.

▸ **7.1** - Computing enhances communication, interaction, and cognition.

▸ **7.4** - Computing innovations influence and are influenced by the economic, social, and cultural contexts in which they are designed and used.

▸ **7.5** - An investigative process is aided by effective organization and selection of resources. Appropriate technologies and tools facilitate the accessing of information and enable the ability to evaluate the credibility of sources.

Lesson 3: Identifying People With Data

Overview

Students begin this lesson by investigating some of the world's biggest data breaches to get a sense for how frequently data breaches happen within companies and organizations, and what kinds of data and information is lost or given up. Afterwards, students will use the Data Privacy Lab tool to investigate just how easily they could be uniquely identified with a few seemingly innocuous pieces of information. At the conclusion of the lesson, students will research themselves online to determine just how much someone could learn about them by conducting the same searches and "connecting the dots."

Purpose

While there are many potential benefits associated with the collection and analysis of large amounts of data, these advances pose a constant risk to our collective security and privacy. Large-scale data breaches mean that the details of our personal, professional, and financial lives may be at risk. In order to prevent personal data from being linked to an individual person, personally identifying information, such as name, address, or identification number, is often removed from publicly available data. Nevertheless, through the use of computational analysis, it is often possible to "re-identify" individuals within data, based on seemingly innocuous information. As more of our lives is digitized, questions of security and privacy become ever more prevalent.

Agenda

Getting Started (10 mins)

- **Explore: World's Biggest Data Breaches Visualization**

Activity (30 mins)

- **Data Privacy Lab: How easily can you be identified? Research Yourself Online**

Wrap-up

- **Class shares their findings**

Assessment

Extended Learning

Objectives

Students will be able to:

- Explain privacy concerns that arise through the mass collection of data
- Use online search tools to find and connect information about a person or topic of interest.
- Explain how multiple sources of data can be combined in order to uncover new knowledge or information.
- Analyze the personal privacy and security concerns that arise with any use of computational systems.

Preparation

- [ ] Familiarize yourself with the external web sites and tools involved in this lesson.

Links

Heads Up! Please make a copy of any documents you plan to share with students.

For the Teacher

- **Activity Guide KEY - Research Yourself** - Answer Key

For the Students

- **Code Studio Unit 4 - Identifying People with Data**
- **World's Biggest Data Breaches Visualization** - Web Site
- **Data Privacy Lab** - Web Site
- **Activity Guide - Research Yourself** - Activity Guide **Make a Copy**

Teaching Guide

Getting Started (10 mins)

Explore: World's Biggest Data Breaches Visualization

Distribute:

Direct students towards the **World's Biggest Data Breaches Visualization - Web Site** (link in code studio) They should spend a couple minutes browsing through the different breaches there. Ask them to make a few notes about the following questions:

- What kind of data is being lost? And how much?
- What kinds of issues could arise from this data getting into the wrong hands?

Discussion

Students will have been thinking primarily about the beneficial effects of collecting and analyzing data. This look at data breaches is intended to be a transition into a set of lessons exploring the potential harmful effects of collecting data, specifically with regards to privacy and security.

Discuss: In small groups or as a class, have students share their findings. The main points to draw out from this conversation are:

- **All kinds of personal data, from usernames to social security numbers and credit card information, is lost fairly regularly.**
- **This information can be used to steal money or identities, get access to classified information, blackmail people, etc.**

Transition:

We've spent a lot of time looking at potential benefits of collecting and analyzing data. As we've already seen today, however, there are some risks associated with collecting all of this information. If it falls into the wrong hands or is used in ways we didn't intend, there may be serious risks imposed on our privacy or security. We're going to start looking more deeply at this problem.

Activity (30 mins)

Data Privacy Lab: How easily can you be identified?

Remarks

> In the data breaches we just looked at, some fairly important pieces of information were stolen. Credit card numbers, passport information, or government security clearances are obviously not something we'd like to fall into the wrong hands. Other pieces of information, however, don't seem that bad. So what if people know your ZIP code? So what if people know your birthday? This is information we usually share without a second thought.

Distribute:

- Direct students to the **Data Privacy Lab - Web Site**.
- Students should type in their information (birthday, ZIP code, and gender) to determine how many other people share those characteristics.
- In most instances, they will find that those three pieces of information can uniquely identify them.

Teaching Tip

If the Data Privacy Lab tool does not work as well for some individuals, they could try the birthday and ZIP code of a parent or close relative.

You should try to keep the Data Privacy Lab tool to 10-15 minutes. Make sure to **keep in mind** the main part of the activity is the second half when students research themselves.

Thinking Prompt:

- **"Why is it significant that you are one of only**

a few people with your birthday, gender, and ZIP code? What concerns does this raise?"

Discuss:

In small groups or as a full class, students should discuss their responses. The main points to draw out from this conversation are:

- We can be uniquely identified from just a few pieces of information.
- Even information we would not normally consider to be "sensitive" can still be used to identify us.
- There are security and privacy concerns raised as a result of most information about us being available online.

Remarks

> As we just saw, there are security and privacy issues that are raised, even when small, seemingly unimportant pieces of information are available online. Most of the time, we don't actually think about what kinds of information are available about us, or how someone might connect the dots with that information.

Research Yourself Online

Activity Guide - Research Yourself

- **Distribute: Activity Guide - Research Yourself - Activity Guide**. Students will work individually and will need access to a computer and the Internet. They will be asked to research themselves online, making note of any and all pieces of information they are able to find. Some guidelines follow:
- They should focus their attention on information that is already publicly available (e.g., through a Google search, on the public pages of their school website, a social network, etc.)
- If students are prevented from accessing some sites on the school's network, they should still list information they know is publicly available elsewhere.
- Students should try to make connections between the data they find. "If I knew this about me and that about me, then I'd also know …"

Teaching Tip

Timing the Activity: Students should be given 15-20 minutes to research themselves, filling in their findings on the Activity Guide. This activity can likely grow or shrink to fit the time you have in your period, but leave time for a wrap-up discussion at the end of the class.

Appeal to the Class Tracker Data from Unit 2: It's possible that such "connecting-the-dots" issues came up when the class was looking at the class tracker data. This would be a good time to bring those issues to the fore. Even though that data was anonymously collected, in aggregate you might be able to identify individuals if you knew even a little bit about them.

Alternate Version: Some students may not have extensive online presences. In these instances, you can ask students to research another member of the community (public official, business person, community leader, etc.) There should still be plenty of fodder for discussion later.

Wrap-up

Class shares their findings

Time permitting it's very interesting to share findings.

Prompt:

- **"What information were you able to find about yourself? Were you able to make connections in the data you collected to figure out anything else? Were you concerned about anything you were able to find?"**

Wrap Up

This activity is aimed at opening the conversation about privacy and security in a highly personal way. The main goal of this closing discussion is just to share what students found about themselves.

Discuss: Students should share their findings, either in small groups or as a class. The main points to draw out from this conversation are below:

- **A great deal of information about us is freely and easily available online.**
- **By making connections in this data or to other sources of data, it is possible to form a more complete picture of who we are and what we do.**
- **There are security and privacy concerns raised by the data we post online about ourselves.**

Teaching Tip

Rather than having this class discussion, you might put students in pairs or small groups. This would reduce both the time needed and potentially sensitive information being over-exposed.

Assessment

- See the examplar response to the worksheet. Available in the teacher only area lesson 3 in code studio **Code Studio Unit 4 - Identifying People with Data**
- See other assessment items in code studio.

Extended Learning

1. You may want to check out **Chapter 2 of Blown to Bits**, which goes into some depth about issues and concerns with data and privacy. In particular, pages 32-35 are related to this lesson.
2. It takes a bit more reading but the Data Privacy Lab project out of Harvard has another fascinating (and scary) project called **The Data Map**
    - You could take a lot of the information there for more rigorous research into how data can be used to identify people

Standards Alignment

Computer Science Principles

- **3.2** - Computing facilitates exploration and the discovery of connections in information.
- **3.3** - There are trade offs when representing information as digital data.
- **7.3** - Computing has a global affect -- both beneficial and harmful -- on people and society.

Lesson 4: The Cost of Free

Overview

This lesson focuses on the economic and consumer concerns around apps and websites that collect and track data about you in exchange for providing you a service free of cost. Often the quality of the service itself is dependent on having access to data about many people and their behavior. The main take-away of the lesson is that students should be more informed consumers of the technology around them. They should be able to explain some of the tradeoffs between maintaining personal privacy and using innovative software free of cost.

Purpose

Many consumers are unaware, or lack a sophisticated understanding, of how information about us is being collected and tracked by the technology we use every day. This issue goes beyond instances when data is stolen from companies or organizations we willingly provide it to. Instead, using computational tools, our movements through the physical and virtual world are being automatically tracked, stored, and analyzed. Cookies in our browsers keep a record of our movements on the Internet. Companies trade access to free tools and apps for the rights to track the data we upload to them. Advertisers develop personalized profiles of potential customers to better target advertising. Governments monitor traffic across the Internet at scales unimaginable without the use of computers. Yet we live in a world that increasingly relies on these digital tools, services and products. Most companies make great efforts to balance the tradeoffs between utility and privacy, but the issues can be tricky and raise confounding ethical dilemmas. We must now grapple with a question of just how much we value our privacy, and whether it is even possible to maintain in a digitized world.

Agenda

Getting Started (15 mins)

Video - The Future of Big Data
Thinking Prompt: What do you know about data collected about you every day?
Quick Poll and Recap of Findings:

Activity 1 (30 mins)

Part 1 - Reading: Wall Street Journal: Users Get as Much as They Give

Activity 2 (30 mins)

Part 2 - Read a real data privacy policy

Objectives

Students will be able to:

- Explain how and why personal data is exchanged for use of free software.
- Explain some of the privacy and economic tradeoffs involved in the collection and use of personal data.
- Describe the ways and reasons organizations collect information about individuals.
- Read and critically evaluate a data privacy policy.

Preparation

- [ ] Review the reading
- [ ] Review the teaching tips related to group work

Links

Heads Up! Please make a copy of any documents you plan to share with students.

For the Teacher

- **EXEMPLAR - Activity Guide - Privacy Policies** - Exemplar

For the Students

- **Activity Guide - Privacy Policies** - Activity Guide **Make a Copy**
- **Code Studio Unit 4 - Lesson 4**
- **The Future of Big Data** - Video
- **WSJ article [original]** - External Article
- **WSJ article [annotated]** - Article **Make a Copy**

Wrap-up

- **Where do you stand?**
- **Would you install this “free app”?**

Assessment

Extended Learning

Teaching Guide

Getting Started (15 mins)

Video - The Future of Big Data

Opening Remarks

> Yesterday we looked at ways that data we willingly give away could be lost and used to compromise our security.
>
> What we often don't think about, however, is just how much data is being collected about us without us even knowing it.
>
> Especially as computers become ever more powerful and ubiquitous, it is becoming easier for vast amounts of data about us to be collected and for it to be used for a variety of purposes.

Video:

- **The Future of Big Data - Video**
- Video is also linked in Code Studio for Students

Transition:

The video mentions how your phone and websites you use track certain things about you. Today we're going to find out a little bit more about it. Here are the primary questions we're interested in:

- Why is this tracking necessary? What are the benefits and drawbacks?
- How can you find out what kind of data is tracked about you and by whom?

Thinking Prompt: What do you know about data collected about you every day?

Prompt:

- **"Write down 2 or 3 websites, web services, or apps that you use the most or rely on the most to stay connected to friends and family, or use for "productivity" like school work."**
- **"For each website / service / app, fill in the following information - just what you know off the top of your head from your own experience or memory"**:

Info to write down for each site:

Students can just complete this in a journal or notebook. They should make a little table in order to compare side-by-side

1. **Name of Website** / Service
2. **Do you have an account** , or need to login?
3. **What kinds of data** does (or could) this site potentially collect about you?
4. Do you know **if this data is shared with other people**, companies or organizations? (If so, which ones?)
5. Do you know **how you would find out** what data is collected or how it's shared?

Give students about 5 minutes to write things down.

Types of sites

Here is a list for you to help jog your students' memory. Encourage students to try to pick 2 or 3 different types, but also ones they've **actually** used.

- **Education:** Code.org, Khan Academy, Codecademy.com
- **Social media:** Facebook, Twitter, Instagram, Snapchat
- **Online store:** Amazon, Target, Walmart
- **Search:** Google, Bing
- **Maps:** MapQuest, Yahoo Maps, Google Maps
- **Productivity:** MS Office Online, Google Docs
- **Mail & communication:** Gmail, Hotmail, Yahoo Mail, Skype, Google Hangouts
- **Streaming sites:** Netflix, Spotify, Pandora
- **Gaming sites:** Steam, Xbox Live
- **Banks and financial institutions:** Chase, Citibank

Quick Poll and Recap of Findings:

- **Show of hands**: How many of the apps that you chose were free?

  This will likely be all or almost all the apps.

- **Whip around**: Name one piece of data the app you chose could potentially collect or know about you.

  Try to get out as many different types as possible.

Transitional Remarks

> - Wow, that's a lot of data! If this stuff is "free," but these companies make a lot of money, then it stands to reason that we are "paying" for these services with our data, because that's the only thing we're giving them in return for a service.
> - What is the monetary value of your personal data?
> - How is it used to make money? What are the tradeoffs? Let's learn a little bit more.

Activity 1 (30 mins)

Part 1 - Reading: Wall Street Journal: Users Get as Much as They Give

Distribute:

Wall Street Journal: **It's Modern Trade: Web Users Get as Much as They Give**

- **Original** article on wsj.com: **WSJ article [original] - External Article**
- **Annotated/Jigsaw** shorter version for jigsaw option: **WSJ article [annotated] - Article**

You can have students read individually or break it up to "jigsaw" it. Either option in the end will probably take the same amount of time. (See teaching tip for more info)

Discuss: Ask students in small groups to discuss what they learned from the reading and come up with a quick "temperature check":

- **"Right now, which way are you leaning? Too little privacy? Right amount?"**
- **"Are you willing to give up some privacy (and potentially some security) to have free access to modern innovative tools - do you trust companies to be good stewards of your data?"**
- **"Are you concerned? Do you think too much of your data is out of your control? Do you think too much personally identifiable data is given over to someone else?"**
- **"What other questions do you have?"**

Goal of Reading WSJ

We want students to understand a bit about how collecting data is part of the business model for many apps / websites and that **we often trade our personal data in order to get these services for free**.

It's slightly complicated by the fact that the services are overall better because of how many people use them while sharing data. But it does raise privacy concerns.

Teaching Tip

Individual reading The full article is about 1200 words. It would probably take about 5-10 minutes for a person to read it. You could read quietly and then put students into small groups for the discussion.

Jigsaw We've prepared an annotated version that breaks up the article into 4 sections. It's recommended that the reading be divided up three ways: Everyone should read the paragraphs denoted "Intro and Background" and then assign each of the remaining 3 sections to groups of students.

Use whatever jigsaw reading strategy you like so that students understand the key ideas of the reading.

Notes for the discussion:

- In the discussion, it's worth noting the source of the article - The Wall Street Journal - and asking students to consider whether or not this information has a "pro-business" slant.

- This is a very complex issue. The important thing is whether or not students are seeing both sides of the issue.
- Try not to let students take the middle ground - “just the right amount of privacy” - as it’s an easy out. Press them to consider a specific case of one app or website and make a determination about that.
- **Caution:** Try to keep the conversation focused on economic terms and the central question of “What is the cost of ‘free’?” It can be easy for this slip into a debate about privacy versus utility, in terms of government access to data, espionage, terrorism, etc. These are extremely important issues as well, but the conversation might get unwieldy. The focus of this lesson is about students becoming more informed consumers of the technology they use.

Activity 2 (30 mins)

Part 2 - Read a real data privacy policy

Transitional Remarks

> When you use most apps, websites, and social networks, **they are collecting information about you in exchange for providing you a service**, like connecting with your friends and sharing photos. Sometimes the service itself, like GPS, needs to track you just to be a useful app.
>
> Other times, the data collected is useful to the company for making money.
>
> Most of the companies that do track your data work hard to balance the tradeoffs between providing you with a service for free and the inherent risks such data collection poses to your personal privacy and security.
>
> But what do they actually collect, and how do they use that data?
>
> Let’s find out.
>
> Most of these companies and organizations (the ethical ones) have a clear, well-written privacy policy. You’re going to pick one to investigate and report back.

Goal of reading a privacy policy

We want students to see that it is possible to look up what data companies are collecting and who they share it with and why.

As an informed consumer, you should at least understand the issues well enough to make informed choices or take action.

Activity Guide - Privacy Policies

Distribute:

- **Activity Guide - Privacy Policies - Activity Guide**

Prompt:

- **"Pick one of the apps / websites that you chose at the beginning of class, and go find and read through that site’s privacy policy"**

Here is a synopsis of what's in the activity guide for students to research.

- Students are asked to note what information the site says they collect, how they are using it, and (hopefully) how they are protecting it.
- The actual activity guide provides a bit more guidance for students about how to find answers to these questions.

> 1. Choose a Website and Find the Data Privacy Policy
> 2. What kinds of data are being collected? How many different kinds of data?
> 3. What service or feature is enabled by the data they are collecting? Why are they collecting it in the first place?
> 4. Who else is given access to that data? How are they using it?
> 5. Can you get access to your own data? Can you modify what is collected or used, or delete your data if you wish?
>
> On a scale of 1-4, rate how comfortable you are with this company’s data policy? 1 - very uncomfortable 4-very comfortable

Discussion / Share out:

- Put students into groups of 3 to share what they found with each other.
- Each group should report out 4 things for the policies reviewed in the group
  1. The names of the companies / organizations / websites reviewed by the group
  2. Notable similarities and differences in the kinds of data collected
  3. Just the number: How many privacy policies let you access, modify or delete your personal data?
  4. Just the number(s): How did you rate the policies on how comfortable you were?
- Teacher should take note of the “comfort ratings.”

Prompt:

- **"What’s your “temperature” on data collection now? Are you leaning toward more privacy? Or the same/less as there is now?**"

Wrap-up

Where do you stand?

Prompt:

- **"This lesson is entitled The Cost of ‘Free.’ What does that mean to you now?"**
- **"How would you explain The Cost of ‘Free’ to a family member, or person you just met, if you had only 60 seconds?"**

Would you install this “free app”?

As a final thought, what is your reaction to this app installation screen? (You can see a higher-resolution version in Code Studio.)

- What questions do you have?
- What would you want to know?
- What would you do to find that out?
- Bottom line: Would you install this app?

Assessment

- To get an idea of what to look for in the activity guides see the **student exemplar** in the teacher only area for Unit 4 lesson 4 here: **EXEMPLAR - Activity Guide - Privacy Policies - Exemplar**

Extended Learning

Articles in the news

Leaning Pro-Utility	Leaning Pro-Privacy
1. **Wall Street Journal: It's Modern Trade: Web Users Get as Much as They Give**	1. **Apple: A Message to Our Customers (Apple challenges order to give government data about terrorist shooter)**
2. **CNN: Despite Facebook, privacy is far from dead**	2. **CNN: The Internet is a surveillance state**

Leaning Pro-Utility	Leaning Pro-Privacy
3. **ZDNet: A case against online privacy**	3. **CNN: Google knows too much about you**
4. **U.S. News: The Case for Internet Surveillance**	4. **TechRepublic: Why "Nothing to Hide" misrepresents online privacy**
5. **Kaspersky: 10 Cool Big Data Projects**	5. **Huffington Post: The Case Against Monitoring Teens Online**
6. **Fortune: Boston is using big data to solve traffic jams**	6. **Politico: We Are All Big Brother Now**
7. **Maclean's: The real reason crime is falling so fast**	
8. **U.S. News: Relax and Learn to Love Big Data**	
9. **LinkedIn: The Ethics of Privacy: The Benefits of Data Gathering**	

Longer Reading

- **Blown to Bits - Chapter 2 - Naked in the Sunlight: Privacy Lost, Privacy Abandoned**
- **Program or Be Programmed** - Chapter 7: Social - Do Not Sell Your Friends

Standards Alignment

Computer Science Principles

- **3.3** - There are trade offs when representing information as digital data.
- **7.3** - Computing has a global affect -- both beneficial and harmful -- on people and society.

Lesson 5: Simple Encryption

Overview

In this lesson, students are introduced to the need for encryption and simple techniques for breaking (or cracking) secret messages. Students try their own hand at cracking a message encoded with the classic Caesar cipher and also a Random Substitution Cipher. Students should become well-acquainted with idea that in an age of powerful computational tools, techniques of encryption will need to be more sophisticated. The most important aspect of this lesson is to understand how and why encryption plays a role in all of our lives every day on the Internet, and that making good encryption is not trivial. Students will get their feet wet with understanding the considerations that must go into making strong encryption in the face of powerful computational tools that can be used to crack it. The need for secrecy when sending bits over the Internet is important for anyone using the Internet.

Purpose

This lesson is the first in a series of lessons about cryptography and encryption. “Encryption” is a process for transforming a message so that the original is “hidden” from anyone who is not the intended recipient. Encryption is not just for the military and spies anymore. We use encryption everyday on the Internet, primarily to conduct commercial transactions, and without it our economy might grind to a halt.

This lesson gives students a first taste of the kind of thinking that goes into encrypting messages in the face of computational tools. Computational tools dramatically increase the strength and complexity of the algorithms we use to encrypt information, but these same tools also increase our ability to crack an encryption. Developing strong encryption relies on knowledge of problems that are “hard” for computers to solve, and using that knowledge to encrypt messages.As a resource, you may wish to read all of Chapter 5 of **Blown to Bits**. It provides social context which you may want to bring to your classroom.

Agenda

Getting Started (15 mins)

- **The critical role of encryption in everyday life**
- **Classic Encryption - The Caesar Cipher**

Activity (35 mins)

- **Part 1 - Crack a Caesar Cipher**
- **Part 2 - Crack a Random Substitution Cipher**

Objectives

Students will be able to:

- Explain why encryption is an important need for everyday life on the Internet.
- Crack a message encrypted with a Caesar cipher using a Caesar Cipher Widget
- Crack a message encrypted with random substitution using Frequency Analysis
- Explain the weaknesses and security flaws of substitution ciphers

Preparation

- [ ] Examine both versions of the widget

Links

Heads Up! Please make a copy of any documents you plan to share with students.

For the Students

- **Caesar Cipher Widget** - Widget
- **Frequency Analysis Widget** - Widget
- **Unit 4 on Code Studio**

Vocabulary

- **Caesar Cipher** - a technique for encryption that shifts the alphabet by some number of characters
- **Cipher** - the generic term for a technique (or algorithm) that performs encryption
- **Cracking encryption** - When you attempt to decode a secret message without knowing all the specifics of the cipher, you are trying to "crack" the encryption.
- **Decryption** - a process that reverses encryption, taking a secret message and reproducing the original plain text
- **Encryption** - a process of encoding messages to keep them secret, so only "authorized" parties can read it.
- **Random Substitution Cipher** - an

Wrap-up
Assessment
Extended Learning

encryption technique that maps each letter of the alphabet to a randomly chosen other letters of the alphabet.

Teaching Guide

Getting Started (15 mins)

The critical role of encryption in everyday life

Thinking Prompt:

- **"In your daily life what things do you or other people rely on keeping a secret? Who are these secrets being kept from? How are these things kept secret?"**

Share:

- Provide a couple minutes for students to share their ideas with their classmates.
- Ask them to brainstorm as many areas as they can where they or other people rely on secrecy.
- Try to touch on as many different people and contexts as possible.

Remarks

> Secrecy is a critical part of our lives, in ways big and small. As our lives increasingly are conducted on the Internet, we want to be sure we can maintain the privacy of our information and control who has access to privileged information.
>
> Digital commerce, business, government operations, and even social networks all rely on our ability to keep information from falling into the wrong hands.
>
> **Recall:** As we saw with our activities on the Internet Simulator the internet is NOT secure
>
> - We need a way to send secret messages...

Discussion

Kick the day off with a **very** quick (5 mins) Think-Pair-Share to activate students' prior knowledge about encryption and why we want it.

Motivate the need for means of encrypting information as it travels across the Internet.

Potential areas to discuss:

- Social interactions (e.g., a surprise birthday party)
- A play in a sports game, your hand in a card game
- Personal identification information, PIN numbers, etc.
- Business and government negotiations
- Military activity

Content Corner

If necessary recall some of the facts we learned in Unit 1 while using the Internet Simulator.

- The Internet is not inherently secure.
- Packets traveling across the Internet move through many routers, each of which could be owned by different people or organizations.
- So we should **assume** all information traveling across the Internet to be public, as if written on a postcard and sent through the mail.

Classic Encryption - The Caesar Cipher

Background:

Many of the ideas we use to keep secrets in the digital age are far older than the Internet. The process of encoding a plain text message in some secret way is called Encryption

For example in Roman times Julius Caesar is reported to have encrypted messages to his soldiers and generals by using a simple alphabetic shift - every character was encrypted by substituting it with a character that was some fixed number of letters away in the alphabet.

As a result an alphabetic shift is often referred to as the Caesar Cipher.

Prompt:

- This message was encrypted using a Caesar Cipher (an "alphabetic shift").
- Let's see how long it takes you to **decode this message** (remember it's just a shifting of the alphabet):

Display or write this on the board

```
serr cvmmn va gur pnsrgrevn
```

- **Give students about 3-5 minutes** to work on cracking the message.
  - **ANSWER: "free pizza in the cafeteria" - the A-Z alphabet is shifted 13 characters.**

Recap:

- With this simple encryption technique it only took a few minutes to decode a small message.
- What if the message were longer BUT you had a computational tool to help you?!

Activity (35 mins)

Cracking Substitution Ciphers

In this set of activities students will use two different versions of a simple widget in Code Studio to "crack" a messages encoded with substitution ciphers, including an alphabetic shift and random substitution.

Transition to Code Studio

Part 1 - Crack a Caesar Cipher

The instructions for this activity are simple - there is no handout:

- Put students in pairs/partners

Goal: Select a message encrypted with a caesar cipher and use the provided widget to "crack" it.

- Experiment with the tool - Click things, poke around, figure out what it's doing.
- **Choose one of the messages from the pull down menu and try to crack it** using the tool.
- If you want to, enter you own message, encrypt it, and have a friend decrypt it.

Give students about 5 minutes to get into the tool and crack a few messages

- Aided with the tool, cracking an alphabetic shift is trivial.
- Once you've done one, it only takes a matter of seconds to do others.

Optional - Pause and Recap:

There is a page in Code studio which recaps terminology (encryption, decryption, crack, cipher, Caesar ciper) and poses the next problem.

You may optionally pause here to recap and go over terms if you like or just let students proceed (see activity part 2 below).

Teaching Tip

Resist the urge to give students a tool or device to aid in cracking this message -- that's coming in the next part of the lesson! Part of the point here is that it's possible without tools. With tools it becomes trivial, as we'll see next.

If students are struggling to start here are a few strategy suggestions:

- Find a small word and try alphabetic shifts until it's clear that it's an English word
- Remember the letters aren't randomly substituted - the alphabet is just shifted.
- Once you have found the amount of shift the rest comes easily.

Content Corner

If you'd like your students to read a little bit about **Historical Cryptography** and cracking ciphers, this lesson and the next one about the Vigenere cipher more or less follow the sequence presented in **Blown to Bits, Chapter 5 - Reading** pp. 165-173.

- Substitution Ciphers and Frequency analysis pp. 165-169
- Secret Keys and One-Time Pads (Vigenere Cipher) pp. 169-173

Teaching Tip

Don't rush it, but don't linger on cracking caesar ciphers. Presenting and cracking a caesar cipher should go **pretty fast**.

The widget is pretty self-explanatory. Let students figure out how to use it on their own.

The goal here is make points about cracking encryption with computational tools, and start to use some common terms.

You should move on to cracking random substitution relatively quickly.

Part 2 - Crack a Random Substitution Cipher

After re-capping the first activity make sure **students understand the following before proceeding:**

- **Cracking a Caesar cipher is easy...trivial with a computational tool like the one we used.**
- **The next step is to make the encryption slightly harder...**

New Challenge:

- **What if instead of shifting the whole alphabet, we mapped every letter of the alphabet to a random different letter of the alphabet? This is called a random substitution cipher.**
- **The new version of the widget you'll see is a more sophisticated version of the encryption tool that shows you lots of different stuff.**
- **But what it does is bit of a mystery!** Let's check it out...

Get Cracking

- Have students click to the next bubble to see the frequency analysis version of the widget. (It should look like the screen shown below)

Goal: let students explore for **5-10** minutes to see if they can discover what the tool is showing them and allowing them to do.

The tasks laid out for students in code studio are:

- Figure out what is going on in this new version of the tool
- What information is being presented to you?
- Figure out what the the tool let's you do
- As usual: you can't break it. So click on things, poke around.
- If you figure it out you might be able to crack a message encoded with random substitution.

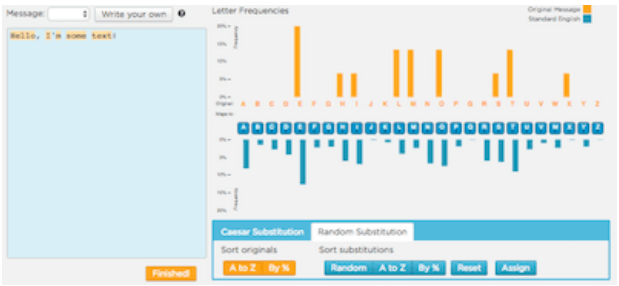
Use a Discovery-based approach

REMINDER: Discovery-based introduction of tools in a nutshell:

- Get students into to the tool without much or any introduction
- Give students working in partners a fixed amount of time (5 minutes or so) to poke around and see if they can figure out what it does and doesn't do – typically this might be presented as a mystery worth investigating
- Ask the group to report what they found
- Teacher fill in any gaps or explanations of how the tool works afterwards

This widget, like all others, are meant as a learning tool. You cannot break it so you are encouraged to let students play and investigate to figure out how the tools work.

These discovery-based methods of introducing tools have been tested in professional development and have worked well for teachers who use this curriculum. This method is effective for a few reasons, but overall students find this approach more engaging and fun, and they tend to be more receptive to, and motivated to hear, explanations of how the tool works after trying to "solve the mystery" themselves.



- **After some exploration time regroup** to clarify what the tool is and how it works.
- **If necessary** point out to students that the next level in code studio (the one after the frequency analysis tool) explains a little bit about how frequency analysis works and suggests a few strategies for how to get started.

Give students about 15-20 minutes to crack one of the messages.

- If they finish there are more to try.

- Students can enter their own messages, do a random substitution to encrypt it, then copy/paste the encrypted version and see if a friend can crack it.
- It is possible to get pretty proficient at cracking these messages with the tool.

Wrap-up

As part of wrap up **the major points we want to draw out are:**

- Encryption is essential for every day life and activity
- The "strength" of encryption is related to how easy it is to crack a message, assuming adversary knows the technique but not the exact "key"
- A random substitution cipher is very crackable by hand though it might take some time, trial and error.
- However, when aided with computational tools, a random substitution cipher can be cracked by a novice in a matter of minutes.
- Simple substitution ciphers give insight into encryption algorithms, but as we've seen fall way short when a potential adversary is aided with computational tools...our understanding must become more sophisticated.
- If we are to create a secure Internet, we will need to develop tools and protocols which can resist the enormous computational power of modern computers.

Here are a couple of thought-provoking prompts you can use to bring closure to the lesson and as an avenue to draw out the points above. Choose one or more.

Prompts:

How much easier is it to crack a caesar cipher than a random substitution cipher? Can you put a number on it?

- **For Caesar's Cipher there are only 25 possible ways to shift the alphabet. Worst case, you only need to try 25 different possibilites. A random substitution cipher has MANY more possibilities (26 factorial = $4x10^{26}$ possibilities). However, as we learned, with frequency analysis we can avoid having to try all of them blindly.**

Was it difficult to crack a Random Substitution cipher? Did it take longer than you thought? shorter? Why?

- **Computational tools aid humans in the implementation of encryption, decryption, and cracking algorithms. In other words, using a computer changes the speed and complexity of the types of encryption we can do, but it also increases our ability to break or circumvent encryption.**

Any encryption cipher is an algorithm **for transforming plaintext into ciphertext. What about the other way around? Can you write out an algorithm for cracking a Ceasar cipher? What about a random substitution cipher?**

- **An algorithm for cracking a Caesar cipher is pretty easy - for each possible alphabetic shift, try it, see if the words come out as english.**
- **An algorithm for cracking random substitution is trickier and more nunanced. There might not be a single great answer but through thinking about it you realize how tricky it is to codify human intelligence and intuition for doing something like frequency analysis into a process that a machine can follow. It probably requires some human intervention which is an interesting point to make.**

Recall that in RFC 3271, “The Internet is for Everyone” Vint Cerf wrote the following. What did he mean by "cryptographic technology?" What does it mean to you now?

Internet is for everyone - but it won't be if its users cannot protect their privacy and the confidentiality of transactions conducted on the network. Let us dedicate ourselves to the proposition that cryptographic technology sufficient to protect privacy from unauthorized disclosure should be freely available, applicable and exportable.

Review of Terminology -- you can use this opportunity to review new vocabulary introduced in the activity and respond to questions students may have encountered during the activity.

- Definitions of cryptography, encryption, decryption, cracking/breaking an encryption, cipher, etc.

Assessment

Questions (also included in Code Studio):

1. What is a Caesar cipher?
2. What is the "key" to a Caesar Cipher that someone needs to know (or discover) to decrypt the message?

   a) A secret word only know by Caesar.

   b) The number of characters to shift each letter in the alphabet.

   c) The letter that occurs most often in the encrypted message.

   d) The day of the month that the encrypted message was sent.
3. The Caesar Cipher has 25 different shifts to try. How many possibilities are there to try in a random substitution cipher?

   a) 26

   b) 26 × 25

   c) 26 × 25 × 24 ×⋯× 3 × 2 x 1

   d) 2626

Extended Learning

Read Blown to Bits

- Read pp. 165-169 of **Blown to Bits, Chapter 5 - Reading**.
- Answer the questions provided in the reading guide and worksheet **Reading Guide for Encryption - Worksheet**
  - For teachers: **Worksheet KEY - Reading Guide for Encryption - Answer Key**

Teaching Tips

Students should be encouraged to chat with their partner while completing the worksheet. The questions are fairly straightforward and the point is more to use the questions as a guide to the reading, than to find all the answers as quickly as possible.

More Blown to Bits

- The earlier sections of Chapter 5 of Blown to Bits make reference to the significance of and controversies surrounding encryption in the aftermath of September 11th. This reading may be a useful tool for further introducing the impact of cryptography on many aspects of modern life.
- Ask students to review the history of their Internet browsing and calculate roughly what percentage they conduct with the assumption that it is "private." Do they have any way of being sure this is the case? Are there any websites they visit where they feel more confident in the secrecy of their traffic than others? Are they justified in this conclusion?

Standards Alignment

CSTA K-12 Computer Science Standards (2011)

▸ **CI** - Community, Global, and Ethical Impacts

- **CL** - Collaboration
- **CPP** - Computing Practice & Programming
- **CT** - Computational Thinking

Computer Science Principles

- **1.2** - Computing enables people to use creative development processes to create computational artifacts for creative expression or to solve a problem.
- **3.3** - There are trade offs when representing information as digital data.
- **6.3** - Cybersecurity is an important concern for the Internet and the systems built on it.
- **7.3** - Computing has a global affect -- both beneficial and harmful -- on people and society.

Lesson 6: Encryption with Keys and Passwords

Overview

In this lesson, students learn about the relationship between cryptographic keys and passwords. Students explore the Vigenère cipher with a widget to examine how a cryptographic "key" can be used to encrypt and decrypt a message. Then, students use a tool that shows them about how long it would take to crack a given password using a standard desktop computer. Students experiment with what makes a good password and answer questions about the “human components” of cybersecurity.

Purpose

Cryptography and encryption are important and far-reaching fields within computer science. This lesson begins to get students’ feet wet with the human side of cybersecurity: choosing good passwords through an exploration of the classic **Vigenère Cipher**. We also learn that the Vigenère cipher is actually susceptible to frequency analysis (though at first glance it is not) and in subsequent lessons we learn that better methods are used today.

Strong encryption techniques are typically publicly known algorithms, but **have mathematical properties** which ensure that the original message cannot easily be retrieved. These techniques typically feature a secret “key” or piece of information that is used when encrypting the message. While the algorithm can be publicly known, the secret key is not. The art of encryption is coming up with an algorithm that 1) makes the message undecipherable without the key and 2) is such that the key should **only** be discoverable through an exhaustive search of all possible keys, rather than through some other analytical technique.

In this lesson we focus on making a good key, while in subsequent lessons we learn more about problems and algorithms that are **computationally hard.** Guessing a random sequence of 200 characters, for example, is computationally hard, because there is no known way to approach the problem besides trying the trillions and trillions of possible character combinations.

Agenda

Getting Started (10 mins)

Think - Pair - Share
Encryption: Algorithms v. Keys

Objectives

Students will be able to:

- Explain the relationship between cryptographic keys and passwords.
- Explain in broad terms what makes a key difficult to “crack.”
- Reason about strong vs. weak passwords using a tool that shows password strength.
- Understand that exponential growth is related to an encryption algorithm’s strength.
- Explain how and why the Vigenère cipher is a stronger form of encryption than plain substitution.
- Explain properties that make for a good key when using the Vigenère Cipher.

Preparation

- [ ] Explore the Vigenere Cipher Widget in Code Studio
- [ ] Familiarize yourself with the "howsecureismypassword.net" site.
- [ ] (Optional) Print out worksheets (links in Code Studio)

Links

Heads Up! Please make a copy of any documents you plan to share with students.

For the Teacher

- **KEY - Exploring the Vigenere Cipher Widget** - Answer Key
- **KEY - Keys and Passwords** - Answer Key

For the Students

- **Exploring the Vigenere Cipher Widget** - Worksheet Make a Copy
- **The Vigenere Cipher** - Widget
- **Keys and Passwords** - Worksheet

Activity 1 (30 mins)

- **Explore the Vigenère Cipher Widget**
- **Recap: Properties of strong encryption**

Activity 2 (20 mins)

- **Computationally Hard Problems -- How good is your password?**

Wrap-up (10-15 mins)

- **Video: Encryption and Public Keys**
- **(Optional) How Not to Get Hacked by Code.org**

Assessment

Extended Learning

Optional Lessons

Make a Copy

- **How Secure Is My Password?** - Code Studio Page
- **The Internet: Encryption & Public Keys** - Video (**download**)
- **(Optional) How Not to Get Hacked** - Resource
- **Unit 4 on Code Studio**

Vocabulary

- **Computationally Hard** - a "hard' problem for a computer is one in which it cannot arrive at a solution in a reasonable amount of time.

Teaching Guide

Getting Started (10 mins)

Remarks

> In the previous lesson you saw how relatively easy it was to crack a substitution cipher with a computational tool.
>
> Today we'll try to crack a different code to see what it's like. Beforehand, however, we should consider why someone might want to crack a cipher in the first place.

Think - Pair - Share

Prompt:

- **"Are there ethical reasons to try to crack secret codes?"**

Give students a few minutes to write down a response and discuss with a neighbor.

Discussion

- Have students quickly share out reasons they came up with.
- There are a lot of different reasons that a person may want to crack a code. Some of them are more ethical (legal) than others.

Discussion

Provide a quick (about 5 minutes) justification for the practice of cracking ciphers, while reviewing relevant vocabulary. At the conclusion of the lesson, students will discuss other reasons we might try to crack a cipher, namely to ensure that it is difficult to do!

Here are some other points that might come out:

- People in the field of **counterterrorism** make a living by trying to crack the codes of other nations. Many attribute the success of the Allies in WWII to our ability to crack the Enigma code and uncover the plans of the Germans.
- Others may try to crack more abstract codes that are not written by humans, searching for **patterns within DNA** models in order to understand their nature and be able to describe the nature of humanity.
- It's useful to try to crack your own codes **to see how strong they really are**.
- There are many other reasons related to mathematical exploration, pattern recognition, etc.

Encryption: Algorithms v. Keys

Today, we will attempt to crack codes, paying particular attention to the processes and algorithms that we use to do so.

So, before starting today we want to make sure that we distinguish between an **encryption algorithm** and an **encryption key**

- An **Encryption algorithm** is some method of doing encryption.
- The **Encryption key** is a specific input that dictates how to apply the method and can also be used to decrypt the message. Some people might say "What is the **key** to unlocking this message?"

Misconception Alert

There's a common misconception that "cracking" and "decrypting" are interchangeable terms.

- **Decrypting** is just using an algorithm to undo the encryption. It's like using a key to unlock a lock. It's what the sender is expecting the intended recipient to do to recover the original message.
- **Cracking** is more like detective work - it's like trying to pick a lock - using various methods to try to figure out what the secret message is without having or knowing the decryption "key" ahead of time.

For example:

- The **Caesar Cipher** is an encryption **algorithm** that involves shifting the alphabet
- The **amount of alphabetic shift** used to encode the message is the **key**
- When you are cracking the Caesar Cipher you are trying to figure out how much the alphabet was shifted - you are trying to discover the key.

Prompt:

- **"If random substitution is an algorithm for encryption, what is the key to a random subsitution cipher?"**
  - A: The key is the actual letter-to-letter mapping that was used to encode the message - it can also be used to decrypt.

Discussion

Quickly review what a “key” is in a cryptographic method and distinguish it from the Algorithm

Transitional Remarks

> So, There is a difference between the algorithm (how to execute the encryption and decryption) and key (the secret piece of information).
>
> - In encryption you should always **assume that your 'enemy' knows the encryption algorithm** and has access to the same tools that you do.
> - What makes encryption REALLY strong is making it hard to guess or crack the “key,” **even if the “enemy” knows** the encryption technique you're using.
>
> Today we’ll learn a little more about it and about keys and their **relationship to passwords** you use every day.

Content Corner

Perhaps counter-intuitively, publicly known encryption algorithms are often more secure, since they have been exposed to a much more rigorous review by the computer science community. Making an encryption algorithm public allows computer scientists to verify the security of the technique either through mathematical proof, or by trying to crack it themselves.

Activity 1 (30 mins)

Explore the Vigenère Cipher Widget

Go to Code Studio

- **Distribute**: **Exploring the Vigenere Cipher Widget - Worksheet**
- Students should click on the **The Vigenere Cipher - Widget**
- Use the worksheet as a guide for exploring the widget.

The **goals** of this activity are:

- Understand how the Vigenère Cipher Algorithm works
- Understand why simple frequency analysis doesn’t work against this cipher
- Figure out what makes for a good v. bad secret key

The activity guide asks students to:

> **Part 1: Explore the Widget**
>
> Students are asked to:
>
> - Jump into the tool and poke around
> - Figure out what it's doing
>
> The worksheet gives a few directed tasks:
>
> - Encrypt a few different messages using different secret keys
> - Decrypt a message
> - Find a “bad” secret key
> - Find a “good” secret key

Teaching Tips

- The Vigenere Cipher Widget is another fun tool to mess around with.
- The key take-aways for students are:
  - A well-chosen key makes a difference - there are certain keys that don’t produce good results.
  - We’re approaching much stronger encryption because we don’t need to keep the encryption method a secret.
  - For example, if I told my enemy that I encrypted a message with the Vigenère cipher, my enemy would still have to do a virtually impossible amount of work to crack the code.
  - Even if I told my enemy the length of the key I used, as long as that length is sufficiently large, it would still leave my enemy basically randomly guessing the key. (Even for this simplified tool, if the key is 10 letters, then there are 26^10 possible keys, ~141 trillion.)
- Try to keep students’ focus on the properties and relationships of the keys to the strength of the encryption.

- Try to decrypt without knowing the key

Part 2: Answer Questions

Students are given space to write answers to these questions.
You can find sample responses in the **KEY - Exploring the Vigenere Cipher Widget - Answer Key**

- Describe in your own words what the Vigenère Cipher Algorithm is doing.
- What makes for a good v. bad secret key using the Vigenère cipher?
- Compare and Contrast the difference between a substitution cipher (Caesar or Random) and Vigenere, using the message “I think I can I think I can I think I can” to explain why Vigenère is a stronger form of encryption than a substitution cipher.
- Will frequency analysis work to crack the Vigenère cipher? Why or why not?
- (paraphrase) Is it easier to crack a message if you know that it was encrypted with the Vigenère Cipher Widget?
- (paraphrase) Is it easier to crack a message if you know that it was encrypted with the Vigenère Cipher Widget **and** that the key was 10 characters long?

Recap: Properties of strong encryption

You may wish to review students' responses on the activity guide at this point. Or you can move that to the wrap-up. We'd like to make a few points about encryption before moving to the next activity...

Prompt:

- **"From what you've seen what are the properties of the Vigenere Cipher that make it harder to crack? In other words, if you had to crack a vigenere cipher what would you do?"**

Discussion

A few points should come out in discussion:

- Vigenere is strong because looking at the cipher text there are no discerable patterns assuming a good key was chosen.
- Because the ciphertext is resistant to analysis it leaves us simply having to guess what the key is.
- Even if we know the length of the key we might still have to try every possible letter combination which is a prohibitively large number of possiblities.

Remarks

- For a long time, the Vigenère cipher was considered to be an unbreakable cipher and was used by governments to send important messages.
- But in the 1800s Vigenere was discovered to be susceptible to a modified form of frequency analysis. After that point it was considered insecure.
- Still the properties of Vigenere that we've found are desirable.

Content Corner

If you are interested in how the Vigenere cipher can be cracked there are a number of resources out there. See the "Extended Learning" section of the lesson plan for links.

Activity 2 (20 mins)

Computationally Hard Problems -- How good is your password?

Introduction

- We know that a good encryption algorithm reduces the problem of cracking it to simply guessing the key.
- We want the key to be **Computationally Hard** to guess - in other words, hard for a computer to guess.

- **Computationally Hard** typically means that arriving at the solution **would take a computer a prohibitively long time** - as in: centuries or eons.
- In terms of cracking encryption that means that the number of possible keys must be so large, that even a computer trying billions of possible keys per second is unlikely to arrive at the correct key in a reasonable amount of time.
- Nowadays when you use a password for a website or device, **your password is used as a cryptographic key**.
- So, choosing a good password is meaningful because we want the key to be hard for a computer to guess. How good is your password?...

Teaching Tip

Don't worry too much about the precise definitions of "computationally hard" and "reasonable time" here. It will be addressed more in the video at the end of this lesson as well as the next lesson.

You should know that the CSP Framework does have a learning objective that relates: **4.2.1 Explain the difference between algorithms that run in a reasonable time and those that do not run in a reasonable time. [P1]**

Go to Code Studio

- Distribute: **Keys and Passwords - Worksheet**
  - The worksheet simply has questions on it to answer. You may distribute them in some other format if you like.
- Students should click on the next page in Code Studio:**How Secure Is My Password? - Code Studio Page** ...

How Secure is my Password - Code Studio Page

Students should read the text on this page about password security and choice.

Student tasks are listed...

1. Open up password strength checker

Students should open the external website **howsecureismypassword.net** in a separate tab or window and then try out these things listed:

2. Test some passwords

Try different passwords to see what the tool tells you:

- Try typing common words from the dictionary or well-known names like “apple” or “chicago”.
- Try typing something that’s over 16 characters.
- Try a string of 4 random words together, like AppleChicagoBalletTree.
- Type a 0. Then keep typing 0s and watch what happens to the statistics. (Actually, you might want to just hold 0 down for a while.)
- Try other things that interest you.

3. Answer Questions

Questions are listed in **Keys and Passwords - Worksheet**:

- Create a few passwords using 8 lowercase ASCII characters (a-z). What’s the longest amount of time-to-crack you can generate?
- Using any characters on the keyboard, what’s the longest amount of time-to-crack you can generate with an 8-character password?
- As you try passwords, what seems to be the single most significant factor in making a password difficult to crack? Why do you think this is?
- Opinion: Is an 8-character minimum a good password length for websites to require? Give your opinion, yes or no, and explain why you think that.
- The AP CS Principles framework contains the following statement: Implementing cybersecurity has software,

Teaching Tips

Make sure you leave enough time for the wrap up.

Students may have a lot of questions about passwords and security that you feel like you might not be able to answer. That’s OK!

a) You don’t have to be an expert on this subject

b) The reality is that the world of cybersecurity changes every day

c) Some of the details can get very complicated, even for professionals.

So, encourage the students’ curiosity and perhaps say, “I don’t know, but I bet you could look it up.” Cybersecurity is an enormous topic. If students get interested, they could dedicate their whole life to this field.

> hardware, and human components. Based on what you've learned so far, describe at least one way that cybersecurity involves "human components."
>
> Hopefully you can now appreciate this comic: **http://xkcd.com/936/**

Wrap-up (10-15 mins)

Discuss:

- Before the Vignere cipher was cracked, many governments openly used it. That is, they made no secret about the fact that they were using the Viginere cipher - it was publicly known. In the modern day, it remains the case that **most encryption techniques are publicly known**.
- **Prompt:** Why might it actually be a good thing that encryption algorithms are freely shared, so that anyone who wishes can try to crack them?
  - **If the security of an encryption technique relies solely on the method remaining a secret, it actually may not be that secure.**
  - **Ideally, a method will be so secure that even if you know which technique was used, it is difficult or impossible to crack the message.**
  - **By making encryption techniques public, we open them up to being tested by anyone who wishes to ensure there are no clever ways of cracking the encryption.**

Discussion

The goal here is to recall that the reason we want to have encrypted transactions is for our own security.

We should feel good about well known strong encryption methods.

We want a world in which anyone can conduct secure transactions on the web; without this possibility, many things would be impossible.

Video: Encryption and Public Keys

- Show the **The Internet: Encryption & Public Keys - Video** (optional)

You should know about this video:

- **0:00 to 4:11** covers Caesar and Vigenere ciphers and explains why they are hard to crack
- **After 4:11**...it explains the difference between encryption that uses **symmetric** v. **asymmetric** keys which is **related to material in the next lesson** and is intended as a preview.
- The next lesson begins by recalling symmetric v. asymmetric keys and getting into how they work.

Wrap up

The video re-iterates a number of points that came out in this lesson.

In wrapping-up, make sure students:

Understand the relationship between cryptographic keys and passwords.

- A **Key** is an input to an encryption algorithm. A password is basically the same thing.

Understand why using longer passwords makes them harder to guess.

- Longer passwords increase the number of possible keys making it **Computationally hard** to guess what the key is.

Transitional Remarks

> We're circling in on some powerful ideas of how secure communication works on the Internet these days. But we need to learn two more things:
>
> 1. We've seen how keys relate to the strength of encryption, but we haven't seen the other side of it -- how modern encryption algorithms actually work. Vigenère was cracked, so what are we using now? In order to do this, we need to understand what kinds of problems are "hard" for computers to solve.
> 2. Right now, the only encryption we know uses a "symmetric key" -- both sender and reciever need to know the secret key, and so they need to meet ahead of time.
>
> But is it possible for you and me to have a secure, private, encrypted exchange without meeting ahead of time and agreeing on a secret password.
>
> The answer is "yes," and we'll find out how it works in the next lesson.

(Optional) How Not to Get Hacked by Code.org

You may want students read or review this little site put together by Code.org

- **How Not To Get Hacked**

Assessment

The worksheet contains several questions for assessment. Here are some additional questions (also included on Code Studio):

1. (Choose two.) Why is the Vigenère cipher hard to crack?

   a) One cannot solve using frequency analysis directly.

   b) Long keys create exponential growth possibilities.

   c) The key is always secret to both the sender and receiver of the message.

   d) A Vigenère cipher relies upon an "alphabet shift" algorithm.

2. What problems exist with encryption schemes such as the Vigenère cipher, even when strong encryption keys are used?
3. Why are computers better than humans at breaking encryptions such as the Vigenère?

   a) Computers are smarter than humans.

   b) Computers are faster than humans.

   c) The Vigenère was originally designed by a computer.

   d) They are not; humans are better as breaking Vigenère encryptions than computers.

4. Which makes for a password that is harder to crack?

   a) A word from the dictionary

   b) 8 random characters that include numbers and punctuation

   c) A 16-character password that is all letters of the alphabet

   d) A 32-character password that is all letters of the alphabet

   e) A 150-character password that is all the same character. ANSWER: E

5. Companies and organizations commonly require users to change their passwords frequently. Websites have password length and complexity requirements. Is it better to change your password frequently or to have a longer password? What level of security is appropriate to require of end users? Does this change, depending on the context (for example, employee or customer)?

Extended Learning

- Go down the rabbit's hole of encryption at **Crypto Corner: http://crypto.interactive-maths.com/**
- - Assign each student a type of cipher. Students should then research the cipher, including information on its algorithm, its history, and what they would have to do to crack the cipher. They should present an example, and describe the process they follow in cracking the code.
- Caesar Cipher video from Khan Academy: https://www.khanacademy.org/computing/computer-science/cryptography/crypt/v/caesar-cipher
- Real world stories of cracking codes:
  - Cracking a code as the key to understanding of ancient culture: http://www.wired.com/2012/11/ff-the-manuscript/all/

- Cracking the human genome (NOVA Video): http://video.pbs.org/video/1841308959/
- Nobel prize given for cracking the code of DNA: http://www.nobelprize.org/educational/medicine/gene-code/history.html
- Navajo Code Talkers http://navajocodetalkers.org/

- Read this quick overview that Code.org put together about **How Not to Get Hacked**, which summarizes some basic cybersecurity issues and how to prevent them.

- Read **Blown to Bits (www.bitsbook.com), Chapter 5**, Secret Bits, pages 161-165, then answer the following questions:
  - The opening pages of Blown to Bits, Chapter 5, discuss a move the government made to try to control encryption in the aftermath of the terrorist attacks of September 11, 2001, but then dropped. Additionally, during the 1990s, the US Government was pressuring the computer industry to be allowed to have a "back door" to decryption. Why do you think they stopped urging for this? http://www.nytimes.com/2013/09/06/us/nsa-foils-much-internet-encryption.html?pagewanted=all&_r=0 (Teacher notes: This would weaken the public's trust in the Internet as an e-commerce vehicle. Any back door could probably be exploited by others. The government believed they could eventually break cryptography without a back door.)
  - Encryption is clearly seen as essential to Internet commercial activity. That it will not be outlawed seems like a settled matter. But conversely, should it be required by government regulation? What about for other non-web media, such as mobile phone traffic and television?
- Vigenère cipher cracker tool on Simon Singh's website. It's a lot of fun and fairly similar to the frequency analysis tool used in class http://www.simonsingh.net/The_Black_Chamber/vigenere_cracking_tool.html
- Have students find videos demonstrating these or other advanced encryption methods; ask them to describe each algorithm and what causes it to be "hard."

Optional Lessons

To dive deeper into the notion of computationally hard problems, consider the following 2 optional lessons after this lesson:

- **Hard Problems - The Traveling Salesperson Problem**
- **One-way Functions - The WiFi Hotspot Problem**

Standards Alignment

CSTA K-12 Computer Science Standards (2011)

- **CPP** - Computing Practice & Programming
- **CT** - Computational Thinking

Computer Science Principles

- **2.3** - Models and simulations use abstraction to generate new understanding and knowledge.
- **3.1** - People use computer programs to process information to gain insight and knowledge.
- **4.2** - Algorithms can solve many but not all computational problems.
- **6.3** - Cybersecurity is an important concern for the Internet and the systems built on it.

Lesson 7: Public Key Cryptography

Overview

This is a big multi-part lesson that introduces the concept of public key cryptography which is an answer to the crucial question: **How can two people send encrypted messages back and forth over insecure channels (the Internet) without meeting ahead of time to agree on a secret key?** In a nutshell, there are two main principles we want students to understand:

1. The mechanics of communication with public key cryptography
2. The basic mathematical principles that make it possible

The lesson gets at these two core ideas through a deliberate chain of thought experiments, demonstrations, activities and widgets. All parts are building blocks that lead to deeper understanding of how it works.

Purpose

This is a fairly hefty lesson because the underlying ideas are subtly quite sophisticated. It's worth noting that much of the material here - all but the highest level takeaways - are beyond the scope of what's covered on the AP exam. Students need to know the basic public key encryption process, and what asymmetric encryption is. For programming they need to know how the modulo operation works.

Our purpose here is to reveal some of the magic that happens every day on the Internet to enable secure transactions. To many the fact that encrypted messages can be sent between parties who have never met before is both taken for granted and opaque. Our belief is that understanding how it works with some depth - getting to experiment with the mathematical principles that make asymmetric keys possible, and the resulting encryption hard to crack - is deeply satisfying.

The widget in the lesson **mimics the RSA encryption algorithm** (with smaller numbers and slightly easier mathematics).

Agenda

Getting Started (5 mins)

- **How do you get the encryption key?**

Activity (15 mins)

- **Step 1: A New Analogy - Cups and Beans**

Activity (30-45 mins)

Objectives

Students will be able to:

- Explain what the modulo operation does and how it operates as a "one-way" function
- Follow an asymmetric encryption algorithm to encrypt a numerical message using the Public Key Crypto widget.
- Explain the difference between symmetric and asymmetric encryption.
- Describe the basic process of encrypting data using public key encryption
- Explain the benefits of public key cryptography

Preparation

This lesson will likely take two days to complete. Preparing for these activities the first time will take some time. Once you've been through it once, the activities actually go quicker than you might expect.

Suggested Prep for Day 1 (Steps 1-3)

- [ ] Prepare the Cups and Beans demonstration (you need cups and beans)
- [ ] Understand the modulo thought experiment with pictures of clocks
- [ ] (Optional) Paper copies of "multiplication + modulo" activity guide

Suggested Prep for Day 2 (Step 4 + wrap up)

- [ ] Practice using the "modulo clock"
- [ ] Practice and Prepare for the using and demonstrating the public key crypto widget

Links

Heads Up! Please make a copy of any documents you plan to share with students.

For the Teacher

- **Public Key Crypto Widget Activity** - Teacher Guide **Make a Copy**
- **Modulo Clock Thought Experiment** -

Teacher Guide **Make a Copy**

- **KEY - Multiplication + Modulo** - Answer Key
- **Public Key Bean Counting (Cups & Beans Activity)** - Teacher Demonstration Guide **Make a Copy**

For the Students

- **(Optional) Public Key Bean Counting** - Activity Guide **Make a Copy**
- **Multiplication + Modulo** - Activity Guide **Make a Copy**
- **CSP Unit 4** - Code Studio
- **The Internet: Encryption & Public Keys** - Video (**download**)
- **(Optional) Public Key Cryptography Recap** - Handout **Make a Copy**
- **How and Why Does the Public Key Crypto Really Work?** - Resource **Make a Copy**

Vocabulary

- **asymmetric encryption** - used in public key encryption, it is scheme in which the key to encrypt data is different from the key to decrypt.
- **modulo** - a mathematical operation that returns the remainder after integer division. Example: 7 MOD 4 = 3
- **Private Key** - In an asymmetric encryption scheme the decryption key is kept private and never shared, so only the intended recipient has the ability to decrypt a message that has been encrypted with a public key.
- **Public Key Encryption** - Used prevalently on the web, it allows for secure messages to be sent between parties without having to agree on, or share, a secret key. It uses an asymmetric encryption scheme in which the encryption key is made public, but the decryption key is kept private.

Teaching Guide

Getting Started (5 mins)

How do you get the encryption key?

Prompt: "How can two people send encrypted messages to each other if they can't communicate, or agree on an encryption key ahead of time, and the only way they have to communicate is over the Internet?"

- You should assume that an adversary is always secretly eavesdropping on their conversation too.
- With a partner come up with a strategy they could use to send encrypted messages.

Discuss

- Give students a few minutes to discuss
- Don't let the discussion go too long
- Direct the conversation toward the idea from the video of using **different keys** - one to encrypt and one to decrypt.

Recall asymmetric keys were mentioned in the cryptography video.

- If you need to show the video **The Internet: Encryption & Public Keys - Video** in whole or in part - the public key cryptography portion starts around the 4:11 mark.

Discussion

Goal: Realize the difficulty of the problem. No form of symmetric encryption will work. There is no way for parties to establish a shared key without agreeing ahead of time in a way that secures it from an observer. Hopefully some students will recall from the video in the last lesson the ideas of using **different keys** - one to encrypt data and one to decrypt it.

Possible Responses: Students may come up with some fantastic ideas, but most will amount to some secret ahead-of-time agreement about a key, or simply some strategy that obscures the key ("security by obscurity").

Transitional Remarks

> Today we're going to dig in a little bit deeper to how this idea of using different keys actually works. The ideas behind how it works are sophisticated, and so to get a deeper understanding we're going to do a series of short activities that stringing together several different ideas, bringing them all together in the end.
>
> Ready? Here we go!

Activity (15 mins)

Step 1: A New Analogy - Cups and Beans

Groups:

- Option 1 (preferred): Teacher Demo. We recommend doing this activity as a teacher demonstration in the interest of time. **Instructions and teacher guide below**

Teaching Tip

Remind students - we're still a ways from the real thing but we're taking baby steps to string ideas together.

- Option 2: Groups of 3 Students. You can have students work through an activity guide that explains it as well. It will take more time. **(Optional) Public Key Bean Counting - Activity Guide**

Materials: Cups and Beans - enough for a demonstration (or for groups of 3, if running as student activity)

Display: You may want to display a picture of a jar full of candies to give a visual for the analogy you're about to explain.

Teacher Demonstration - Cups and Beans

The lock box analogy from the video is a good start, but our first step to seeing how public key cryptography works requires us to look at the same process of using public and private keys but with an analogy that goes a step further.

Full Teacher Guide: Public Key Bean Counting (Cups & Beans Activity) - Teacher Demonstration Guide

Setup and Activity Summary:

Discussion

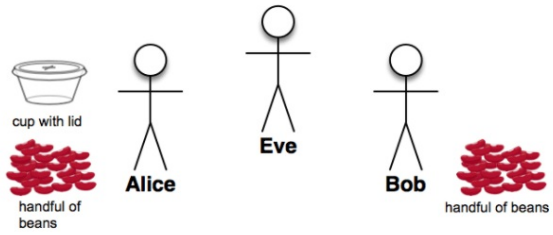
The cups and beans demo showed **basically the same** public/private key analogy as the lockbox in the video.

Similarities:

- For Bob to send a message to Alice he needs to obtain a public key, which we can use to "lock" a message
- Only Alice can "unlock" the message
- Bob and Alice do not need to agree on a key ahead of time
- Alice never lets her private key out in public

Differences:

- Beans in cups is closer to how **data** is encrypted - beans are data, sealed in the jar is encrypted
- Eve (or anyone else) could only guess what was in the jar even though it passed right in front of/through them over the "Internet"
- At no point was the secret message ever out in public, or sent unsecured.
- Closer to reality: Notice how the public key itself is a form of encrypted message. But it's used to encrypt **something else**



- Alice choose a private key (some number of beans)
- Alice make a "public key cup" by placing beans in a clear cup and sealing it
- Pass the cup to Bob over the "Internet"
- Bob grab the "public key cup" and add a secret number (of beans) to it
- Pass the cup back to Alice over the "Internet"
- Alice open cup and subtract the number of beans she added originally
- What's left is Bob's secret number

Discuss: Relate this process using cups and beans to the lockbox analogy from the video. What's similar? What's different? What took place of the public key? The message? The private key?

- Let students discuss for a minute
- You may review the "what's the point" items and table at the end of teacher demonstration guide
- Ensure that students see how the cups and beans process was similar to the lockbox process.

Remarks

> Okay so that's one step. We now have a clearer idea of the public key encryption**process**. If we can keep extending this we'll have a solution to the problem of how two people can encrypt messages without meeting ahead of time.
>
> **Next** we need to see how actual data is encrypted rather than beans in cups.
>
> To learn that, we'll need to string a few more ideas together.

Activity (30-45 mins)

Step 2: Modulo - The operation behind public key encryption

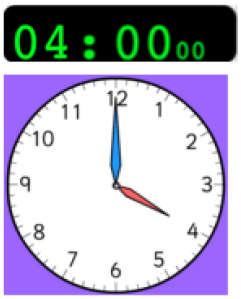
The next idea we need to add is an important mathematical operation called "modulo".

Remarks

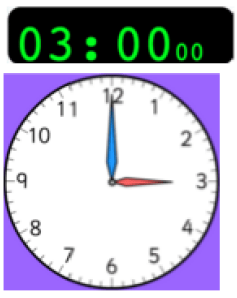
> The cups and beans demonstration showed us how the mechanics of public key cryptography works.
>
> It's a big deal that asymmetric encryption allows for two parties to send secret messages to each other over public channels without having to agree on a secret encryption key ahead of time.
>
> Now let's look at the mathematical principles that allow private and public keys to work.

"Clock Arithmetic Thought Experiment"

> **Teacher Guide:** Use the **Modulo Clock Thought Experiment - Teacher Guide**
>
> Here is a summary:
>
> **Materials**: two pictures of analog clocks - one with hour hand at 4:00 and another at 3:00.
>
> **Display**: picture of clock at 4:00. You can use this **interactive clock** rather than pictures if you like.



BEFORE: clock is showing 4:00.



AFTER: clock is showing 3:00.

> **Run the thought experiment** : Use Full Teacher Guide for details: **Modulo Clock Thought Experiment - Teacher Guide**
>
> Key Points of the thought experiment:
>
> - This "clock" operation is called **Modulo**
> - Modulo is an actual math operation - it's the remainder after division
> - The clock is a useful visual to think about, but the size of the clock is arbitrary
> - the same principle of "wrapping" around the clock would apply no matter how many ticks were on the clock.

What's Modulo?

The modulo operation is a math operation that returns the **remainder** from dividing two numbers. For example, in classic divison 13/5 is **2 Remainder 3** . The mod operation gives the remainder portion. So we would say `13 MOD 5 = 3` .

There is a well known visual analogy for modular arithmetic using clocks since modulo is often thought to "wrap" the number system. If, for example, you use 12 as a modulus then **any** result must be in the range 0-11 since those are the only possible remainders. Similarly, no matter how many hours you count off on a traditional analog clock, there is a limited number of hours (1-12) that the hour hand can be pointing to. It's even called "Clock Arithmetic" in some places **wikipedia: modular arithmetic**

The **modulo operation is important for cryptography** because it can act as a one-way function - the output obscures the input.

No need to

The purpose of this thought experiement is to understand the clock analogy for modulo. It is a setup for the next step.

Students should understand the concept of numbers that "wrap" around the clock and that the "size" of the clock could be arbitrary - it doesn't have to be 12. The same principle would apply for a "clock" of any size.

Step 3: The Mod Clock Widget and Multiplication + Modulo

Remarks

> Modulo is important for cryptography as a **one way function** - you can't tell based on the remainder what went into the clock.
>
> To understand **how** it's used in cryptography, we're going to investigate what happens when we use **simple multiplication** to produce the number we input into the clock. There are certain properties that are useful when we **combine** simple multiplication with modulo.

Multiplication + Modulo Activity

Group: Have students partner up in groups of 2 or 3

Distribute: Activity guide **Multiplication + Modulo - Activity Guide**

Code Studio: Direct students to the "Mod Clock Widget" in code studio.

- Demonstrate a few quick sample inputs to show how the clock size can change and numbers "wrap around"
- The big number in the middle is the remainder, the result of the modulo operation

Student do the activity: students should work with a partner to work through the problems on the activity guide.

Circulate as students work. Make sure that they are trying out the problems given which ask them to try to guess numbers. They should also be using the Mod Clock to check their results.

Students should get a feel for this general formula: `(A * B) MOD M` and its properties, because it is the foundation on which we'll create public and private keys in the next step.

Discuss: **"Why is it hard to guess which numbers multiplied together produce the result?"**

These points are made at the bottom of the activity guide. After students have worked on the problems for a bit they should be able to give a few responses here such as:

- You cannot solve it like an equation in math class

Mod Clock + Multiplication

This step has two goals:

1. Allow students to play with the "Mod Clock" widget to get a sense for how modulo works
2. See how multiplication combined with modulo can lead to "computationally hard" problems to solve

In particular we want students get a feel for how and why guessing the blank value is pretty hard in: `A * ____ MOD M = R` even when you know A, M, and R.

For example, guess the missing value in this: `47 * ____ MOD 51 = 1` you are essentially reduced to random guessing.

This is not on the AP

Students **do not** need to memorize or be facile with these mathematics for the AP Exam.

The modulo operation is part of the AP pseudocode and there might be simple programming questions on the exam that use it.

However, the mathematics for Public Key Cryptography is beyond the scope of the course. We are giving it a small treatment here to expose a statement from the AP CSP framework: **6.3.1I Cryptography has a mathematical foundation.***

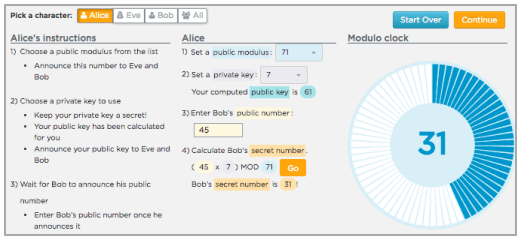
Content Corner

You cannot solve it like a typical equation in math class because there are many equations. If you are looking for `A * ___ MOD 13 = 1` for example, what you are really trying to find is a number that you could multiply by A that comes out to one of a list of infinite values: 1, 14, 27, 40, 53,...and so on.

- Numbers kind of jump all over the place
- You kind of have to just guess randomly, or at least systematically try every number.

Activity (30-45 mins)

Step 4: Use the Public Key Crypto Widget Activity



The public key crypto widget showing Alice's screen

Activity

- Use the widget to practice the public key encryption process
- Explain how asymmetric encryption works at a high level
- See how multiplication + modulo can be used to create asymmetric keys
- Try to crack messages encrypted with multiplication+modulo

Bringing it home

> Okay, now to finally bring everything together. This is last and final step in which we'll see how we can use the math we just learned about to create public and private keys.

The Public Key Crypto Widget Activity

Teacher Guide: Use the **Public Key Crypto Widget Activity - Teacher Guide** which contains details for each step of this process.

Group: Put students into groups of 2 (to play just Alice and Bob initially).

- Each student should be at their own computer, but within speaking distance

Display: the Public Key Crypto Widget Instructions page (in code studio)

- You can ask students to go to that page as well if you want them to read it now, or just have it displayed for you to review the instructions.

Summary: Use the teacher guide, but here is a summary for reference:

Real Public Key Cryptography?

It might be hard to believe but this widget is pretty close to mimicking real RSA encryption.

When you use RSA "for real" you have to generate a public/private key pair using software on your computer. You put the public key somewhere that someone can grab, like your personal web page (there are other ways too.)

You keep the private key on your computer and never distribute it.

Most of the time your computer handles the encryption and decryption behind the scenes.

If you would like to try or demonstrate for your students, you can. Just google "RSA Keygen" and follow instructions for your type of computer.

Part 1: Introduce the widget (10 mins)

- Introduce the Public Key Crypto widget providing the background and instructions given on the Instructions page in code studio. Make sure to point out the similarities and differences between using this widget and cups and beans.
- Demonstrate the first step of using the widget. (Click past the the instructions page to get to the widget if necessary)

Part 2: Just play Alice and Bob (5 mins)

- With a partner, just play Alice and Bob and exchange a few numbers to get the hang of it. Communicate by just speaking out loud. Exchange roles at least once. Verify that you can encrypt and decrypt messages.

Part 3: Show how Eve works (10 mins)

- After pairs have gotten the hang of playing Bob and Alice, regroup to review how Eve works. Display Eve's screen in the widget.
- Pick 2 students on opposite sides of the room to play Alice and Bob and demonstrate intercepting their spoken broadcasts and entering the info in Eve's screen.

Part 4: Experiment with cracking bigger numbers (5-10 mins)

Note: Grouping Options

- Option 1: Crowd-source cracking - Continue as a whole class, with 2 students playing Bob and Alice, and everyone else playing Eve.
- Option 2: Small group experimentation - Have previous Alice-and-Bob pairs get together in groups of 4. One pair plays Bob and Alice, the other pair plays Eve as a team of 2 (on one computer or two)
- Students exchange numbers a few more times, trying to make it hard for Eve to crack. See how long it takes and what makes it hard. At what point would you feel "safe" as Alice or Bob that your messages were basically secure? As you play with the widget can you figure out why it works? **Why can Alice decrypt the message but Eve can't?**

💡 Answers to some FAQs about the

Clock size is chosen randomly by Alice but there is a set list of values to choose from. The clock sizes in the list provided are prime numbers between 1 and 10,000. This ensures certain properties of the encryption.

Alice's private key is also chosen at "random" but there is also a list to choose from. We've computed pairs of public/private keys behind the scenes so they have the necessary mathematical relationship. Alice simply has to pick one.

Bob is sending a secret number to Alice , not vice-versa. In public key cryptography for Bob to send a secret to Alice, alice has to act first, producing a public key for Bob to use.

Bob can send any number to Alice - as long as the number is between 0 and (clockSize - 1.)

The clock size limits the range of values - the secret numbers that Bob and Alice use are confined to the output range of the mod clock. For example: if the clock size is 13, then Bob can only send a secret number in the range 0-12. If the clock size is 253 then the secret values can be 0-252.

(Optional) Part 5 - Use the "show all 3" version of the widget

- Look at the "all" tab in the widget, which lets you act out and see all 3 characters at the same time by yourself. Try this out for a few rounds and see if you get a sense for why it works. Encourage students here to play with small values so the can get a sense of the relationships between the numbers.

Optional Recap Handout: There is an optional student handout that Recaps important ideas from the widget:
(Optional) Public Key Cryptography Recap - Handout

Discussion (10 mins)

Discuss: What made the encryption harder/easier for Eve to crack?

- Perhaps obvious, but the bigger the clock size the harder it is for Eve to crack.
- There are also certain values that Bob could send, like 0 or 1, that would give away the secret.
- **There is no way to crack the encryption other than brute force**
- If you could imagine that value being not a 4-digit number but, say a 75-digit number the computation for Eve becomes mind bogglingly hard.

Discuss: Let's problem solve! The widget right now only lets you send one secret number at a time. Furthermore, it's kind of slow - it requires multiple trips over the internet to send one message. What's the fastest way you could use this tool (or any public key encryption) to send a secure text message?

Give students a moment to discuss and brainstorm.

- Students will likely suggest using ASCII codes in some fashion - perhaps trying to cluster more than one ASCII character per message sent.
- Note that if you're going to send multiple messages using public key cryptography you should change the public key occasionally, otherwise you're giving Eve more clues to crack the message with - you want Eve to start over every time.

- A **really clever thing to do** is to only send one number that represents a key both parties can use for a good old fashioned symmetric encryption. In other words, only use (the slower, multi-trip) public key cryptography for the purpose of establishing a secret key to use in some other encryption method.
- This is, in fact how HTTPS works - it uses public key cryptography to establish a secret key between two parties. Once established it uses a much faster encryption method for sending everything else.

Optional Discussion: **According to the widget look at what Eve has to compute to crack Alice's private key. This reveals how Alice's public key was computed based on her choice of clock size and private key. Why are these made so that** `pvt * pub MOD clock = 1` **?**

- The **only** thing students really need to takeaway from this is that Alice's public key is no accident. It was computed to make the math in the end work out. That's all they need to know.
- But, this fact - that the result of Alice's initial computation is `1` - is the crux of why the math works out in the end.
  - Short version: when Alice multiplies bob's encrypted message by her private key, it cancels out the public key portion of Bob's multiplication (because `pvt * pub MOD clock = 1` it's just multiplying bob's number by 1), leaving only Bob's number remaining.
  - You can read a more thorough explanation here:**How and Why Does the Public Key Crypto Really Work? - Resource**

Remarks

> This is as far as we're going to take the public key analogy. The public key crypto widget is a superficial version of RSA encryption. Instead of basic multiplication, RSA:
>
> - Uses numbers raised to powers of large prime numbers
> - Very large (256-bit) values for the modulo divisor (clock size)
> - Crack the encryption requires finding the prime factors of EXTREMELY large numbers. Prime factorization is much harder computational problem to solve than our little multiplication+mod problems here.
>
> But from these activities hopefully you have a better sense of how public key encryption works and how making asymmetric keys is at least mathematically **possible.**

Wrap-up (10 mins)

Why this is important

Remarks

> **Public Key Encryption was (and is) considered a major breakthrough in computer science.**
>
> - Public key cryptography is what makes secure transactions on the Internet possible.
> - In the history of the Internet, the creation of public key cryptography is one of the most significant innovations; without it we could not do much of what we take for granted today --we couldn't buy things, communicate without being spied on, use banks, or keep our own conduct on the Internet secret or private.
> - Until asymmetric encryption was invented, the only way to ensure secure transactions on the Internet was to establish a shared private key, or to use a third party to guarantee security.
> - The implications of this are huge. It means any person can send any other person a secret message transmitting information over insecure channels!

Prompt: We just spent a lot of time learning about Public Key Cryptography through a bunch of different analogies, tools and activities. And what you've been exposed to mimics the **real thing** pretty closely. **But what are the essential elements? Let's do a brain dump! List out what you think are the most important or crucial elements of Public Key Cryptography that you've learned.**

Give students a few minutes to jot down their lists.

Pair & Share: Have students share their lists with an elbow partner. Then share to the whole group. Many valid points and ideas may emerge. Here are the key ones:

1. Public Key Cryptography is a form of asymmetric encryption
2. For Bob to send Alice a message, Bob must obtain Alice's public key
3. The underlying mathematics ensure that both the public key and a message encrypted with the public key are **computationally hard to crack** while making it easy to decrypt with a private key
4. It is strong because the **method** of encryption is publicly known, but keys are never exchanged.

Whittle it down

Goal: We want to ensure that we whittle down all of the various parts of this lesson and distill the things that are really important.

A lot of the activities, analogies and tools were in service of getting to some deep ideas about encryption and how it works. Ultimately, exposure to those deep ideas is helpful, but the actual facts that students need to know about Public Key Encryption are few.

There are some more detailed ideas about Public Key Cryptography that are interesting but not crucial for the AP Exam.

- A public and private key are mathematically related so that decrypting is easy
- The modulo operation acts as a one-way function to obscure inputs that are very large numbers
- No one owns it - it's a public standard

Optional: Make a table applying terminology to the various analogies we saw

Fill in a table that shows all of the terms we've learned around public key encryption and how each analogy we've seen applies.

	Lockbox	Cups & Beans	Public Key Crypto Widget
private key			
public key			
encrypted message			
how to decrypt			
how to crack			

Assessment

Questions:

1. In symmetric encryption, the same key is used to encrypt and decrypt a message. In asymmetric encryption different keys are used to encrypt and decrypt. Give at least one reason (more are welcome) why asymmetric encryption is useful.
2. In the cups and beans activity, what is the public key? What is the private key? What is the unencrypted and encrypted message?
3. What are some other examples of one-way functions? Can you think of a one-way function in real life?
4. Using your name and the name of a friend, describe the process of sending your friend a message using public key cryptography. Your explanation should include the terms: **Public Key, Private Key, Encrypt(ion), Decrypt(ion)**
5. Explain what the modulo operation does. You may use the analogy of a clock in your answer if you like.

6. Why is modulo a one-way function?
7. Describe to a person who knows nothing about encryption why public key encryption is hard to crack.
8. What is 13 MOD 17?

   a) 0

   b) 1 4/13

   c) 4

   d) 13

   e) 17
9. What is 20 MOD 15?

   a) 0

   b) 1.5

   c) 5

   d) 15

   e) 20

Extended Learning

- The Public Key Crypto Widget simulates the basic mechanics of RSA Encryption, with slightly more simple math. You could go read about **RSA Encryption**.
- **RSA Encryption Examples**

Standards Alignment

CSTA K-12 Computer Science Standards (2011)

- **CPP** - Computing Practice & Programming
- **CT** - Computational Thinking

Computer Science Principles

- **4.2** - Algorithms can solve many but not all computational problems.
- **6.3** - Cybersecurity is an important concern for the Internet and the systems built on it.

Lesson 8: Rapid Research - Cybercrime

Research

Overview

Students learn about various types of cybercrimes and the cybersecurity measures that can help prevent them. Then students perform a Rapid Research project investigating a particular cybercrime event with a particular focus on the data that was lost or stolen and the concerns that arise as a result. The Rapid Research activity features vocabulary, concepts, and skills that should help prepare them for the AP Explore PT, and also serves as a capstone for the sequence of lessons on encryption and security.

Purpose

This lesson serves two roles. 1. Review terminology about cybersecurity and crime that is relevant for the AP CS Principles Exam and 2. Practice research and writing skills that will help students on the Explore PT.

Following this lesson you may opt to either run the research activity in the next lesson or move on to running the full Explore PT with your class. Note that the Explore PT prep unit includes additional resources that will help students prepare for the task.

Agenda

Getting Started (5 mins)

- **Video: Cybersecurity & Crime**

Activity (40 + 30 mins)

- **Rapid Reseach - Cybersecurity and Crime**
- **Day 1 - Choose Innovation, Read and Research**
- **Day 2 - Prepare one-pager**

Wrap Up (10 mins)

- **Review Cybersecurity Terms**

Assessment

Objectives

Students will be able to:

- Explain the characteristics of a phishing attack
- Explain how a DDoS attack works
- Describe how one computer virus works
- Research and describe a cyber attack found in the news
- Reason about the threats posed by, and methods of recourse for, various types of cyber attacks
- Describe plausible storage, security, or privacy concerns for particular pieces of data

Preparation

- [ ] Review the video
- [ ] Review annotated responses to terminology at end of lesson for wrap up

Links

Heads Up! Please make a copy of any documents you plan to share with students.

For the Teacher

- **Video Guide KEY for "Cybersecurity and Crime"** - Answer Key

For the Students

- **Rapid Research - Cybercrime** - Activity Guide **Make a Copy**
- **Cybersecurity One-Pager** - Template **Make a Copy**
- **Cybersecurity and Crime Video Worksheet (Optional)** - Video Worksheet **Make a Copy**
- **How Not To Get Hacked** - Web Resource
- **The Internet: Cybersecurity and Crime** - Video (**download**)

- **[Deprecated] Guide: Rapid Research - Cybersecurity and Crime** - Activity Guide **Make a Copy**

Vocabulary

- **Antivirus Software** - usually keeps big lists of known viruses and scans your computer looking for the virus programs in order to get rid of them.
- **DDoS Attack** - Distributed Denial of Service Attack. Typically a virus installed on many computers (thousands) activate at the same time and flood a target with traffic to the point the server becomes overwhelmed.
- **Firewall** - software that runs on servers (often routers) that only allows traffic through according to some set of security rules.
- **Phishing Scam** - a thief trying to trick you into sending them sensitive information. Typically these include emails about system updates asking you send your username and password, social security number or other things.
- **SSL/TLS** - Secure Sockets layer / Transport Layer Security - An encryption layer of HTTP that uses public key cryptography to establish a secure connection.
- **Virus** - a program that runs on a computer to do something the owner of the computer does not intend.

Teaching Guide

Getting Started (5 mins)

Video: Cybersecurity & Crime

Remarks

> To conclude our thinking about encryption and security we're going to look at how cybercrimes are conducted, how cybersecurity measures can protect us, and what the implications are of data leaking. Then you'll research a particular cybercrime and quickly prepare a one-pager about it.

Show: The Internet: Cybersecurity and Crime - Video

- Have students watch the video (display for all, or have students watch in Code Studio)
- Have students complete the **Cybersecurity and Crime Video Worksheet (Optional) - Video Worksheet**

Teaching Tip

NOTE: this video is also embeded at the top of the **How Not To Get Hacked - Web Resource** page that students read in the actvity, as well as in Code Studio. You might choose to send them directly to that at this point or show the video to the whole class.

Content Corner

The video touches on types of cybercrimes and cyber attacks NOT covered in the **How Not To Get Hacked - Web Resource** article but are still vocabulary that students need to know, specifically DDoS attacks and how they work.

The video touches on a number of topics that students might choose to research later:

- DDoS Attacks (and Bot Nets)
- Cyber warfare
- Viruses and Anti Virus Software
- Phishing Scams
- Credit Card theft
- Types of people who commit cybercrimes

Activity (40 + 30 mins)

Rapid Reseach - Cybersecurity and Crime

Distribute: Give students copies of **Rapid Research - Cybercrime - Activity Guide** and **Cybersecurity One-Pager - Template** .

Below is a suggested schedule for completing the project.

Day 1 - Choose Innovation, Read and Research

Review Activity Guide and Rubric: At the beginning of the project, emphasize the importance of reviewing the **one-pager template** and **rubric**. Students may assume that more is required of them than is actually the case. Point out that the written component is quite short. They probably have space for at most 100-150 words per response.

Teaching Tip

Differences from the actual Explore PT: The actual Explore Performance Task will be completed over 8 class hours. The fact that this schedule is significantly shorter reflects several differences in this Practice PT.

- Some categories and topics have been supplied ahead of time.
- Students are not creating any kind of computational artifact
- Students are not describing the beneficial or harmful effects of an innovation / event.

Choosing Your Cybercrime Event: It is recommended that you place a time limit on this process (e.g. 20 minutes). Students should not leave class after the first day without a topic in mind and ideally with some resources identified. Luckily, in choosing their topics, students will likely have begun to identify resources they can use in completing their project.

Conducting Your Research: This document is intended to serve primarily as a guide to students for identifying online sources of information. The skill students need to develop is identifying useful resources on their own and then synthesizing this information. Being presented with a structured way of doing this means students will have a model for how to complete their research when completing the actual Explore PT.

Teaching Tip

Cybercrime Definition: The definition of a cybercrime event as "any instance where digitally stored data falls into the hands of someone not originally intended to have access to it" is used to help align this task to the Explore PT. In particular this definition sets up the last two prompts of the activity guide where students must both specifically identify the data used by an app and describe concerns specifically related to this data. These are critical skills students must use when describing the computing innovation they will research. Make sure you reinforce this definition as students choose their topics.

Day 2 - Prepare one-pager

Complete One-Pager: Students should find this aspect of their project most familiar. The prompts are similar in style and content to prompts students have already seen. Emphasize the need for clarity in their writing, and remind them that everything must fit on a single page. If they have responded completely to each of the prompts, it is fine to write less.

Sharing/Submission: You may want to collect students' one-pagers, have them share in small groups, or with the whole class. Since students were researching something of their own choosing, they might be eager to show what they found out.

Wrap Up (10 mins)

Review Cybersecurity Terms

Below is the list of cybersecurity terms that students were introduced to throughout this lesson.

We've annotated them with brief explanations that should come out during discussion.

Content Corner

These terms are pulled directly from the AP CSP Framework. Check out the mappings to the framework at the bottom of this lesson plan.

These statements can be used as the basis for question on the AP CSP Exam.

The annotations given here should provide enough depth for the kinds of responses expected of students.

Teaching Tip

If you are running out of time, assigning some of these terms for homework might be a good way to review and kick off the next day.

- **Implementing cybersecurity has software, hardware, and human components.**
  - This is a theme for the whole lesson
  - Vulnerabilities in hardware and software can be compromised as part of an attack.
  - But, as mentioned in the video, a large percentage of cybersecurity vulnerabilities are human-related, such as choosing bad passwords, (unintentionally) installing viruses, or giving personal information away.
- **Sockets layer/transport layer security (SSL/TLS)**
  - An encryption layer of HTTP. When you see the little lock icon and `https` it means that you are visiting a website over HTTP but the data going back and forth bewtween you and the server is encrypted.
  - SSL (secure sockets layer) and TLS (transport layer security) usepublic key cryptography to establish a secure

connection.

- **Cyber warfare and cyber crime have widespread and potentially devastating effects.**
  - This is especially true in the case of warfare which (fortunately) we have not experienced much of on a global scale. But using cyber attacks to cripple basic infrastructure (power, water) and communication could be devastating.
- **Distributed denial of service attacks (DDoS)**
  - Typically a virus installed on many computers (thousands) activate at the same time and flood a target with traffic to the point the server becomes overwhelmed -- doing this can render web services like DNS, or routers, or certain websites useless and unresponsive.
- **Phishing scams**
  - Typically a thief trying to trick you into sending them sensitive information. Typically these include emails about system updates asking you send your username and password, social security number or other things.
  - More sophisticated scams can make websites and email look very similar to the real thing.
- **Viruses / Antivirus software and firewalls**
  - A virus is program that runs on a computer to do something the owner of the computer does not intend. Viruses can be used as a Bot Net to trigger a DDoS-style attack, or they can spy on your computer activity, such as capturing all the keystrokes you make at the computer, or websites you visit, etc.
  - Antivirus software usually keeps big lists of known viruses and scans your computer looking for the virus programs in order to get rid of them.
  - A "firewall" is simply software that runs on servers (often routers) that only allows traffic through according to some set of security rules.

Assessment

Rapid Research: Use the rubric provided with the Activity Guide to assess the one-pagers.

Video: These questions refer to ideas in the Cybercrime video.

- What does the `s` in `https` refer to?
  - It's the plural of `http` - a more robust version of http that runs on multiple channels.
  - **s is for "secure" - a version of http that is encrypted.**
  - s is for "simple" - a simplified version of http that runs faster on modern computers
  - s is for "standard" - to distinguish the original http from non-standard versions like `httpv` and `httpx`
- When someone tries to get you to give up personal information through email or a bogus website it is called a:
  - DDoS Attack
  - **Phishing Scam**
  - Virus
  - SSL/TLS layer
- When someone attemps to compromise a target by flooding it with requests from multiple systems that is called a:
  - **DDoS Attack**
  - Phishing Scam
  - Virus
  - SSL/TLS layer
- The vast majority of computer security failures are due to:
  - Software vulnerabilities
  - Hardware limitations
  - **Human carelessness**
  - Bot Nets

Standards Alignment

Computer Science Principles

- **6.2** - Characteristics of the Internet influence the systems built on it.
- **6.3** - Cybersecurity is an important concern for the Internet and the systems built on it.
- **7.3** - Computing has a global affect -- both beneficial and harmful -- on people and society.

Lesson 9: Practice PT - Big Data and Cybersecurity Dilemmas

Overview

To conclude their study of big data and cryptography, students will complete a small research project related to a dilemma presented by Big Data or Cybersecurity, in the form of a Practice Performance Task. Students will pick one of two issues to research more deeply - either an issue related to big data, or one related to cybersecurity. Students will need to identify appropriate online resources to learn about the functionality, context, and impact of the technological innovation that gave rise to the dilemma they are investigating. After completing their research, students will present their findings both in a written summary and with an audio / visual artifact they found online. The written components students must complete are similar to those students will see in the AP Performance Tasks.

This project is an opportunity to practice many of the skills students will use when completing the Explore Performance Task on the AP® Exam at the end of the year. While an open-ended research project might be intimidating, students have built all the skills they need to complete this task.

Note: This is NOT the official AP® Performance Task that will be submitted as part of the Advanced Placement exam; it is a practice activity intended to prepare students for some portions of their individual performance at a later time.

Note for 2017-18 School Year: This Practice PT has NOT been updated to reflect changes to the **Explore PT Scoring Guidelines** released in Fall 2017. We recommend you review those guidelines to understand the similarities between this project and the actual Explore PT.

Purpose

This lesson does not cover new CS content per se, though students might discover new and interesting things in their research. This lesson is an opportunity for students to synthesize their knowledge and understanding of Big Data, cybersecurity, cryptography, and computationally hard problems. The project asks students to tie their research into a topic in the news with vocabulary and concepts covered in this unit of study. For reference, vocabulary and topics from lessons in this unit include:

- Big Data
- Moore's Law

Objectives

Students will be able to:

- Identify reliable and authoritative sources of information about a computing information.
- Synthesize information taken from multiple online sources to create a cohesive description of a computing innovation.
- Identify an artifact that clarifies an aspect of a computing topic not easily captured in writing.
- Explain both the beneficial and harmful effects related to a modern social dilemma in computing

Preparation

- [ ] Review the Practice PT

Links

Heads Up! Please make a copy of any documents you plan to share with students.

For the Students

- **Big Data and Cybersecurity Dilemmas** - Practice PT **Make a Copy**
- **Unit 4 on Code Studio**

- Encryption and Decryption
- Symmetric v. Asymmetric Encryption
- Computationally Hard Problems
- Public Key Encryption

Agenda

Getting Started

Distribute and Review the Project

Activity

Complete the Practice PT

Assessment

Use the project rubric

Extended Learning

Wrap-up

Teaching Guide

Getting Started

Remarks

> At the end of the year you will need to complete the Explore Performance Task. The project we're about to do asks you to conduct research on a big data or cybersecurity dilemma and present your findings both visually and in writing.

Thus, to conclude our study of Big Data and Cybersecurity you will be completing a practice Performance Task on a topic of your choosing. Hopefully this will be an enjoyable opportunity to dig deeper on a topic that piqued your interest over the last few weeks, and it will of course be useful preparation for the Explore Performance Task, which you'll do at the end of the year.

Teaching Tip

Consider Skipping: Depending on how much time you have remaining in the year, you may opt to skip this lesson and move on directly to completing the actual Explore PT, using the AP: Explore PT Prep unit as a guide. If students have demonstrated strong research and writing skills in the Rapid Research lessons (2 and 8) then you'll likely be fine to move on.

Caution - Not Updated to Match 2018 Scoring Guidelines: This Practice PT has NOT been updated to reflect changes to the **Explore PT Scoring Guidelines** released in Fall 2017. The differences are fairly subtle and the task description itself has not changed, just the guidelines used to score it. As such this lesson remains useful practice of the core skills needed to complete the Explore PT and can be a valuable exercise for classrooms that need more practice. You should review the guidelines or the resources in the AP: Explore PT Prep unit if you'd like to better understand the nuances of the task.

Distribute and Review the Project

Distribute: Big Data and Cybersecurity Dilemmas - Practice PT and as a class review the project guidelines and rubric. Respond to questions.

Activity

A proposed schedule of the steps of this project is included below, as well as more thorough explanations of how to conduct the various stages.

Day 1

- Review Project Guidelines and Rubric
- Select a big data or cybersecurity dilemma to research
- Identify online sources of information using the Research Guide

Day 2

- Continue to record findings in the Research Guide
- Identify potential artifacts to include
- Begin writing written responses

Day 3

- Complete any remaining research to answer questions
- Select and make any necessary edits to artifacts
- Complete written responses

Complete the Practice PT

Read Requirements:

At the beginning of the project emphasize the importance of reviewing the rubric. Students may assume that more is required of them than is actually the case. In particular emphasize that they do not need to create their artifact themselves, but it must still meet the requirements of the prompt. Point out that the written component of the project is similiar to what students did for the Practice PT - The Internet and Society in Unit 1.

Choosing Topics:

It is recommended that you place a time limit on this process (e.g. 20 minutes). Students should not leave class after the first day without a topic in mind and ideally with some resources identified. Luckily, in choosing their topics students will likely have begun to identify resources they can use in completing their project.

Teaching Tips

Difference from the actual Explore PT: The actual Explore Performance Task will be completed over 8 class hours. The fact that this schedule is significantly shorter reflects several differences in this Practice PT.

- We have provided topics to focus research
- Several written responses have been eliminated
- Students do not need to create their own artifacts (though they may if they so choose)

The primary goal of this Practice PT is to familiarize students with the format of the Explore PT and the thinking practices they will need to employ when completing it.

Complete the Research Guide:

This document is intended to serve primarily as a guide to students. The skill students need to develop is identifying useful and **credible** resources on their own and then synthesizing this information. Being presented with a structured way of doing this means students will have a model for how to complete their research when completing the actual Explore PT.

Identify an Artifact:

This is perhaps the greatest deviation from the real AP Explore PT. For this, students do not need to create their own artifact. Instead they need to identify an audio or visual artifact (image, visualization, drawing, chart, video, interview, etc.) that highlights a harm or benefit caused by the innovation, or helps to explain it better. This may still be a challenging process. The goal is to help students think about what good audio / visual artifacts look like and how they present complex material. You can recall what students learned from the "Good and Bad Visualizations" lesson from Unit 2. In Unit 2 they also developed skills for developing good computational artifacts on their own.

Written Responses:

Students should find this aspect of their project most familiar. The prompts are similar in style and content to prompts students have already seen. Emphasize the need for clarity in their writing, and remind them that while the 300 word limit is a **maximum** -- they do not necessarily need to write 300 words for each prompt. If they have responded completely to each of the prompts it is fine to write less.

Submission:

For the AP Explore Performance Task students are asked to compile all of their written work into a single PDF. You will need to determine how best to collect this work in your class but you may optionally wish to practice this process when collecting submissions for this project.

Assessment

Use the project rubric

Included in the Practice PT is a Rubric by which the project can be assessed.

Extended Learning

Ask students to look at the beneficial and harmful effects of cybersecurity issues like the NSA spying on emails. The more current and relevant the issue, the better.

The book **Blown to Bits** has several chapters that relate to personal security, privacy and liberty in face of big data and encryption.

Wrap-up

Presentation (Optional): If time allows students may wish to have an opportunity to share their research with one another. Consider other options like creating a “Digital Museum” by posting links to all their projects to a single shared document.

Standards Alignment

CSTA K-12 Computer Science Standards (2011)

- **CD** - Computers & Communication Devices
- **CI** - Community, Global, and Ethical Impacts
- **CL** - Collaboration
- **CPP** - Computing Practice & Programming

Computer Science Principles

- **1.1** - Creative development can be an essential process for creating computational artifacts.
- **1.2** - Computing enables people to use creative development processes to create computational artifacts for creative expression or to solve a problem.
- **6.3** - Cybersecurity is an important concern for the Internet and the systems built on it.
- **7.3** - Computing has a global affect -- both beneficial and harmful -- on people and society.
- **7.4** - Computing innovations influence and are influenced by the economic, social, and cultural contexts in which they are designed and used.