Impact of Computing

Computing Innovations

A **computing innovation** includes a program as an integral part of its function.

A **computing innovation** can be physical (e.g., self-driving car), nonphysical computing software (e.g., picture editing software), or a nonphysical computing concept (e.g., e-commerce).

The purpose of computing innovations is to solve problems or to pursue interests through creative expression. An understanding of the purpose of a computing innovation provides developers with an improved ability to develop that computing innovation.

A **program** is a collection of program statements that performs a specific task when run by a computer. A program is often referred to as **software**. It can be written in different programming languages like Python or Java.

Effects of Computing Innovations

Not every effect of a computing innovation is anticipated in advance. Some effects are harmful.

For example, inventors of television, computers, mobile devices, and social media did not intend to unleash a slew of negative consequences for children.

- shortened attention spans
- lack of connection to nature
- bullying and hate groups
- using computing innovations as tools for deploying fake news

A single effect can be viewed as both beneficial and harmful by different people, or even by the same person. For example, video game makers will disagree about the "lack of connection to nature" argument above.

Effects of Computing Innovations

Many effects of innovations are beneficial.

Advances in computing have generated and increased creativity in other fields, such as medicine, engineering, communications, and the arts.

Computing innovations can be used in ways that their creators had not originally intended:

- The World Wide Web was originally intended only for rapid and easy exchange of information within the scientific community.
- Targeted advertising is used to help businesses, but it can be misused at both individual and aggregate levels, compromising, for example, personal privacy.

Effects of Computing Innovations

Computing innovations can be used in ways that their creators had not originally intended(continued):

- Machine learning and data mining have enabled innovation in medicine, business, and science, but information discovered in this way has also been used to discriminate against groups of individuals.
    - For example, historical data on employment may show women getting promoted less than men. If a machine learning system trained on such data concludes that women are worse hires, it will perpetuate discrimination.

It is not possible for a programmer to consider all the ways a computing innovation can be used.

Digital Divide

Internet access varies between socioeconomic, geographic, and demographic characteristics, as well as between countries.

The **digital divide** is the gap that exists between individuals who have access to modern information and communication technology and those who lack access. This difference in access to computing devices and technology is based on socioeconomic, geographic, or demographic characteristics.

The digital divide:

- can affect both groups and individuals
- raises issues of equity, access, and influence, both globally and locally.
- is affected by the actions of individuals, organizations, and governments.

Computing Bias

Computing innovations can reflect existing human biases because of biases written into the algorithms or biases in the data used by the innovation.

- for example, in association bias, data for a machine learning model reinforces and/or multiplies a cultural bias. Your dataset may have a collection of jobs in which all men are doctors and all women are nurses.

Programmers should take action to reduce bias in algorithms used for computing innovations as a way of combating existing human biases.

Biases can be embedded at all levels of software development.

Crowdsourcing

Crowdsourcing is a sourcing model in which individuals or organizations obtain goods and services, including ideas and finances, from a large group of internet users.

- it divides work between participants to achieve a cumulative result.
- E.g. “idea competitions” and “innovation contests”. (Netflix Prize, Lego Ideas)
- Tedious "microtasks" performed in parallel by large, paid crowds (e.g. Amazon Mechanical Turk, Bitcoin miners) are another form of crowdsourcing.
- Amazon Mechanical Turk (MTurk) is a crowdsourcing marketplace that makes it easier for individuals and businesses to outsource their processes and jobs to a distributed workforce who can perform these tasks virtually.

Crowdsourcing offers new models for collaboration, such as connecting businesses or social causes with funding.

Citizen Science

Science has been impacted by using scale and “**citizen science**” to solve scientific problems using home computers in scientific research

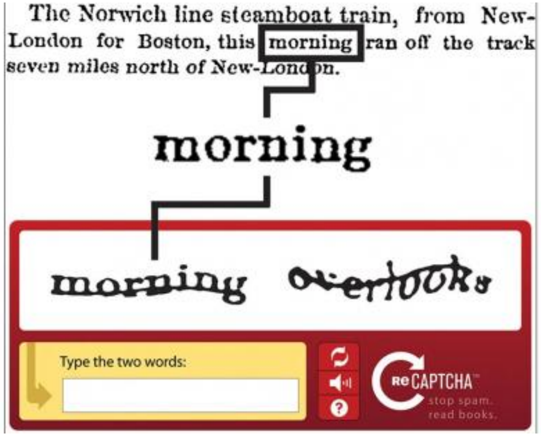
- **Citizen science** is scientific research conducted in whole or part by distributed individuals, many of whom may not be scientists, who contribute relevant data to research using their own computing devices e.g. folding@home(protein folding) and Galaxy Zoo(classify galaxies).
- Some online services use the contributions of many people to benefit both individuals and society, for example, ImageNet which uses Amazon Mechanical Turk to find/hire people to classify/label images to be used for image classification algorithms.
- Human capabilities are enhanced by digitally enabled collaboration.

CAPTCHA VS reCAPTCHA

CAPTCHA(Completely Automated Public Turing test to tell Computers and Humans Apart) is the human validation test (usually the blurry squiglly letters that need to be deciphered) used by many sites to prevent spam.

reCAPTCHA is a reversed CAPTCHA - the same test, used not only to prevent spam but to help in the book digitizing project. In other words, the reCAPTCHA tests are not meaningless combination of words, but excerpts from books that undergo digitation, while CAPTCHA uses several human validation methods including math or general knowledge questions, visual puzzles and even chess puzzles.

- reCAPTCHA has completed digitizing the archives of The New York Times and books from Google Books, as of 2011 and digitized books that are too illegible to be scanned by computers in 2015.



The Norwich line steamboat train, from New-London for Boston, this morning ran off the track seven miles north of New-London.
morning
morning overlooks
Type the two words:
reCAPTCHA
stop spam.
read books.

The Impact of Computing

Computing enhances communication, interaction, and cognition.

- Email, texting(SMS), and video conferencing and video chat have fostered new ways to communicate and collaborate.
- **cloud computing**: performing calculations and modeling on servers that have more resources(AWS(Amazon Web Services), Google Apps, Cloud 9).
- e-commerce(online shopping e.g., Amazon), health care, access to information and entertainment, and online learning.

Widespread access to information facilitates the identification of problems, development of solutions, and dissemination of results.

- Public data, such as databases of temperature readings or databases of court cases, provides widespread access and enables solutions to identified problems.
- Trends of what people search for in the Internet are predictors of behavior.
- Social media, including blogs and twitter, have enabled dissemination.
- Smart grids(electricity):an electricity supply network that uses digital communications technology to detect and react to local changes in usage.

Legal and Ethical Concerns

Material created on a computer is the intellectual property of the creator or an organization.

Ease of access and distribution of digitized information raises intellectual property concerns regarding ownership, value, and use.

Measures should be taken to safeguard intellectual property.

The use of material created by someone else without permission and presented as one’s own is plagiarism and may have legal consequences.

Legal and Ethical Concerns

Open Access and **Creative Commons** have enabled broad access to digital information. Open and curated scientific databases have benefited scientific researchers.

Creative Commons—a public copyright license that enables the free distribution of an otherwise copyrighted work. This is used when the content creator wants to give others the right to share, use, and build upon the work they have created.

cc creative commons

open source—programs that are made freely available and may be redistributed and modified(examples: Firefox browser, OpenOffice(in competition with Microsoft Office).

open access—online research output free of any and all restrictions on access and free of many restrictions on use, such as copyright or license restrictions

Open Access

Global Effects

Computing has global effects – both beneficial and harmful – on people and society.

Innovations enabled by computing raise legal and ethical concerns.

- Downloading movies/music, streaming movies, access to digital content through peer-to-peer networks(for example, BitTorrent)
- the existence of computing devices that collect and analyze data by continuously monitoring activities
- Digital access to digital books(PDF, EPUB)
- Commercial and governmental censorship of digital information
- **Open source** and licensing of software and content.
  - Is human knowledge advanced by full and free access to all information, allowing engineers and developers to correct and improve on already existing systems?
  - Or does a lack of strong protection for **Intellectual Property**(IP) discourage innovation by removing the financial incentive for developing it?
  - What is the balance of these potential benefits over these potential harms?
  - See: https://www.scu.edu/ethics/focus-areas/internet-ethics/resources/unavoidable-ethical-questions-about-open-source/

Privacy

Privacy and security concerns arise in the development and use of computational systems and artifacts.

- **privacy:** Privacy relates to any rights you have to control your personal information and how it’s used. A bank selling your info to marketers without your consent is a breach in privacy.
- **security:** Security, on the other hand, refers to how your personal information is protected. Cybercriminals breaking into the bank's servers and stealing your information is a breach in security.
- Aggregation of information including geo-location, **cookies**, and browsing history raises privacy and security concerns.
  - cookies can track browsing habits and used for ads.

- Targeted advertising is used to help individuals

but it can be misused at both individual and

aggregate levels.

Privacy

Personally identifiable information (PII) is information about an individual that identifies, links, relates, or describes them. Examples of PII include: Social Security number, age, race, phone number(s), medical information, financial information and biometric data.

PII can be used to stalk or steal the identity of a person or to aid in the planning of other criminal acts.

Information placed online can be used in ways that were not intended and that may have a harmful impact. For example, an email message may be forwarded, tweets can be retweeted, and social media posts can be viewed by potential employers.

Proxy Server

Anonymity in online interactions can be enabled through the use of online anonymity software and **proxy servers**.

The idea behind proxy servers is similar to that behind NAT(Network Address Translation) and private IP addresses.

Remember, when you make a request from a private IP address, the server is tricked into thinking that the request actually came from another device, which is the router.

With a proxy server, you can essentially do the same thing: after connecting to a proxy, it can make requests to other web pages on your behalf and then forward you the responses.

That way, only the proxy knows what your IP address actually is, and the websites you're browsing only see the IP address of the proxy.

Cybersecurity

Cybersecurity: set of techniques to protect the **secrecy**, **integrity** and **availability** of computer systems and data against threats.

- Secrecy or confidentiality: only authorized people should be able to access or read specific computer systems and data.
  - E.g. data breaches, where hackers reveal credit card information, is an attack on secrecy
- Integrity: only authorized people should be able to modify or use systems and data.
  - Hackers who learn your email password and then send emails masquerading as you is an attack on integrity.
- Availability: authorized people should always have access to their systems and data.
  - **Denial of Service(DoS)** attacks: Hackers overload a server with fake requests rendering the server slow or unaccessible. This is an attack on the service's availability.

Common Attacks

Phishing attacks: a hacker impersonate either a legitimate person or a corporation through an email that asks the user to take an action that would give the phisher an access point to critical data or information.

- Phisher spoofs the logo or website of a well-known corporation or individual so their email request appears legitimate.

- Contact HR department, posing as a trusted source, to get info(W2, SS#, etc..)

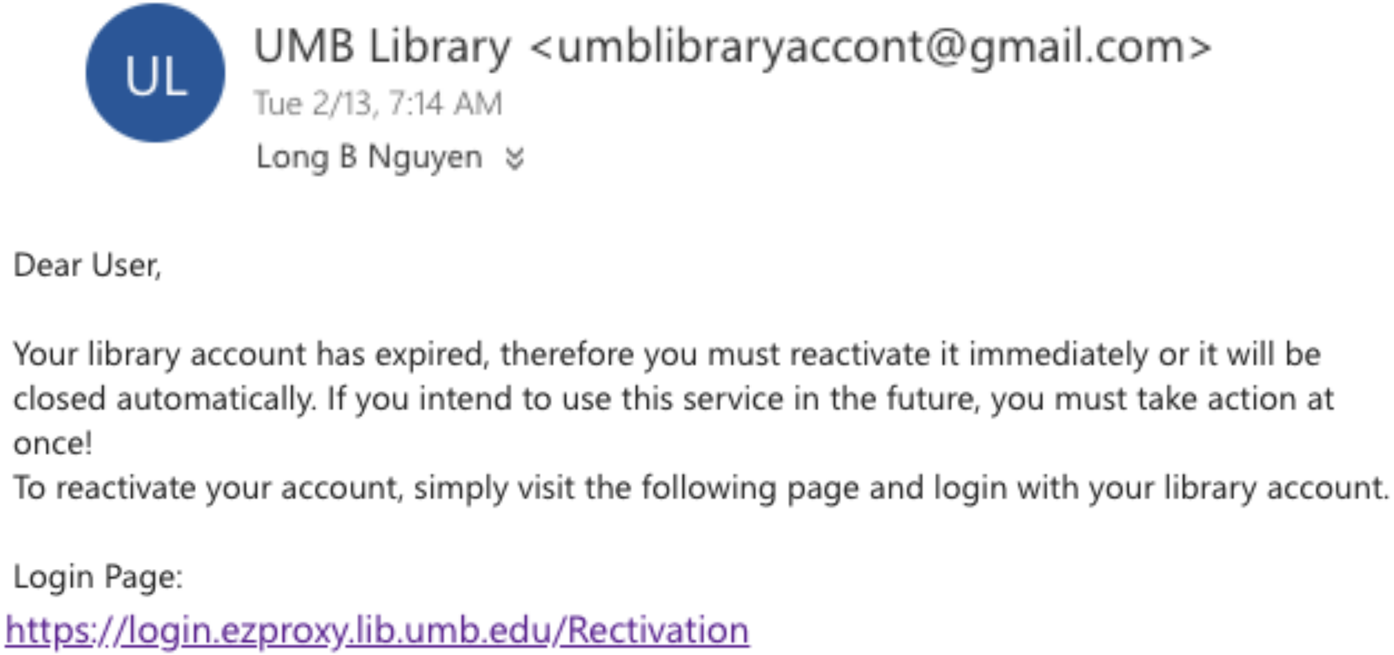
To avoid a phishing attack:

- Pay attention to anything that may be slightly wrong with an email, including misspellings, strange syntax, or logos that have been slightly altered.
- Never to click on a link within an email. For example, if an employee is contacted by their bank and encouraged to reset a password, it’s best to go directly to the bank’s website.

Keylogging is the use of a program to record every keystroke made by a computer user in order to gain fraudulent access to passwords and other confidential information.

A Personal Example of Phishing

Can you spot indicators that this is a phishing email?



UMB Library <umblibraryaccont@gmail.com>
Tue 2/13, 7:14 AM
Long B Nguyen

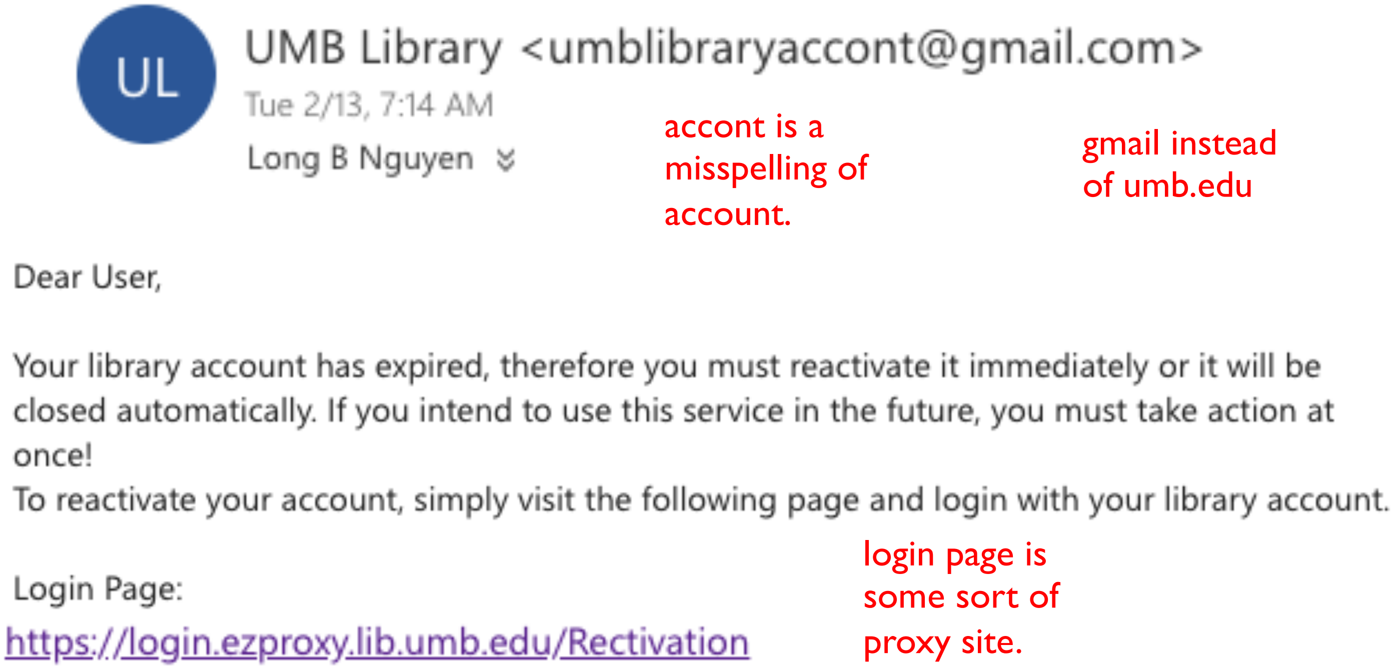
Dear User,

Your library account has expired, therefore you must reactivate it immediately or it will be closed automatically. If you intend to use this service in the future, you must take action at once!
To reactivate your account, simply visit the following page and login with your library account.

Login Page:
https://login.ezproxy.lib.umb.edu/Rectivation

A Personal Example of Phishing

Can you spot indicators that this is a phishing email?



UMB Library <umblibraryaccont@gmail.com>

Tue 2/13, 7:14 AM

Long B Nguyen

Dear User,

Your library account has expired, therefore you must reactivate it immediately or it will be closed automatically. If you intend to use this service in the future, you must take action at once!
To reactivate your account, simply visit the following page and login with your library account.

Login Page:
https://login.ezproxy.lib.umb.edu/Rectivation

accont is a misspelling of account.

gmail instead of umb.edu

login page is some sort of proxy site.

misspelling

Security

Security boils down to two questions:

- Who are you?
- What should you have access to?

Access should be given to authorized people and refused to the wrong people.

To differentiate between authorized/unauthorized personnel, we use **authentication,** the process by which the computer understand who it’s interacting with.

Authentication measures protect devices and information from unauthorized access. Examples of authentication measures include strong passwords and multifactor authentication.

Authentication

Multifactor authentication is a method of computer access control in which a user is only granted access after successfully presenting several separate pieces of evidence to an authentication mechanism, typically in at least two of the following categories:

- Knowledge(what you know): based on a secret knowledge known only to the real user and the computer. For example, username and password.
- Possession(what you have): based on the possession of a secret token that only the real user has. For example, a physical key and lock.
- Inherence(what you are): based on YOU! You authenticate by presenting yourself to the computer.
  - **Biometric authenticators**: e.g. fingerprint readers and iris scanners.

What You Know

Username/Password

- Easy to implement, can be compromised if hackers can guess or find out.
- Some are easy for humans to figure out: e.g. 1234 or 1111
- Computers can try to **brute force** all password combinations. For example, a 4-digit ATM pin only as 10,000 combinations. 0000 to 9999.
  - Some systems will lock you out after 3 failed attempts.
  - But what if hackers have taken over many thousands of computers, trying a random pin like 1056 even once on thousands of computers can gain access to one or more bank accounts.
  - A famous successful brute force attack = The Fappening(See slide #24)
- Many websites now requires upper/lower case, numbers and special symbols to explode the number of combinations.
  - An 8-digit pin has 100,000,000 combinations. But an 8-character password has 600 trillion combinations.

Nand Mirroring

One ineffective type of attack is brute forcing login credentials, e.g. try all combinations of passwords.

most systems defend against this type of attack by locking the user out after a certain number of tries.

Nand mirroring is one hack around this.

- attach wires to memory chip and make a perfect copy of its contents.
- try some password combinations, if the system is locked up, reflash the memory with original copy thereby resetting it and allowing the hacker to try new passwords immediately.
- nand mirroring was effective, for example, on an IPhone 5C.

What You have

What you have: based on the possession of a secret token that only the real user has. For example, a physical key and lock.

- You can open the door if you have the key.
- Avoid the problem of being guessable.
- Harder for remote attacker
- But can be compromised if hacker is close by: keys can be copied, phones can be stolen, even locks can be picked.

What You are

- what you are: based on YOU! You authenticate by presenting yourself to the computer.
    - Biometric authenticators: e.g. fingerprint readers and iris scanners.
    - secure but can be expensive.
    - data over sensors varies over time.

- what you know and what you have methods are **deterministic**
    - **deterministic**: always predicting the same output from a given input.
    - if you know the password or have the key, you're granted access 100% of the time, if you don't, you're granted access 0% of the time.

What You are

- biometric authenticators are **probabilistic**
    - **probabilistic**: element of chance is involved, different output may come from the same input.
    - if lighting is bad or you’re wearing glasses, there’s a chance that the system won’t recognize you.
    - Or worse, it can recognize someone who is not you! (Your twin)

- biometric authenticators can’t be reset
    - what if your hacker compromises your fingerprint?
    - it’s possible to forge an iris by capturing a photo

Hackers

white hats: hackers hired to evaluate systems security, find bugs and security holes in software.

black hats: malicious hackers who intend to steal, exploit, sell private data.

- Blackhat(2015); Chris Hemsworth.

(not sure if I can believe Thor is a hacker.)

why hackers hack:

- for curiosity and amusement
- for monetary gain(**cybercriminals**)
- promote a political or social goal(**hacktivists**)

How Hackers Hack

most common way hackers get into a computer system isn’t by hacking; it’s by tricking people into letting them in.

social engineering: a person is tricked into revealing confidential information.

- most common type of attack is phishing. For example, an email from Bank of America asking to reset the password by clicking on a link to a Bank of America clone site. Login credentials is compromised if account holders attempt to log on.
- even if less than 1% of success rate, a million phishing emails can possibly yield thousands of accounts.
- **pretexting**: an attacker called a company and pretend to be from the IT department and convince a user to configure their computer in a compromising way.

Malware

emails can be a common delivery mechanism for **trojan horses**, programs that masquerade as harmless attachments, like a photo, but actually contain malicious software, called **malware**.

malware can take on many forms.

- some steal confidential data(**spyware**)
  - install keylogger, that records all your keystrokes and send them to a remote source.
- others encrypt data and demand a ransom(**ransomware**)
- **virus**: attach itself to a file or program and harm your computer if opened.
  - delete files, slow down computer
  - requires human interaction(running or opening the file)
- **worms**: similar to virus but can self-propagate without user interaction.

Homework

Watch or Read:

a)Crowdsourcing and Crowdfunding:

https://www.youtube.com/watch?v=-38uPkyH9vI

https://www.youtube.com/watch?v=Buyub6vIG3Q

b) reCAPTCHA:

https://thehustle.co/the-genius-whos-tricking-the-world-into-doing-his-work-recaptcha (Please read!)

https://www.youtube.com/watch?v=PQ-xzwj_p_4 (described by inventor Luis von Ahn himself)

c) Citizen Science

https://www.youtube.com/watch?v=SZwJzB-yMrU

References

1) AP CollegeBoard. AP Computer Science Course Description. (2017).

2) Unversity of Rhode Island. AP Computer Science Principles. Global Impact. Retrieved from

http://computing-concepts.cs.uri.edu/index.php/Main_Page