IT2023 Lecture Diary (Nov 03, 2023)

1 Announcements

1. As I will be in a conference trip from November 18. 2023 to November 25, 2023, 'November 24 (Friday) lecture' is re-scheduled to November 27, 2023 (Monday) evening.
2. The course homepage has been modified.
   - Two links have been added in the end of the *Outline 2*. One is linked to a video about how to use free AI tools to generate an artificial Internet celebrity. The other is linked to a youtube channel introducing the applications of free AI tools in generating contents to be appeared in a youtube channel.
   - A new outline, *Outline 4: Technologies for Information/Intelligent Infrastructure*, has been added with links to the telecommunication technologies.
3. New contents have been added in the document 'Evolution of Technology', including (i) the 'Introduction' section, (ii) Section 4.5.5 and (iii) Section 4.5.7.
4. In the *Lecture Diary* for October 27, 2023, new content has been added in Section 3. Moreover, two new sections have been added, including Section 4 and Section 5.

2 Intelligent Infrastructure

In the document *Introduction to Intelligent Technology*, the idea of *intelligent infrastructure* (version September 23, 2023) has been introduced, see Figure 4 in the document. *Intelligent infrastructure* includes all the AI tools to support the development and delivery of *intelligent applications* (resp. *intelligent services*). Below lists a number of popular intelligent services for use in content generation.

- Image generation: Midjourney.
- Text content generation: ChatGPT, Google Bard and Hugging Face.
- Translation: Google Translate.
- Search sentence recommendation: Google Search and Bing.
- Voice-To-Text generation: Cortona, Google Assistant and Siri.

Note that *he above intelligent application systems play a role as an assistant.* They assist a user to generate content. They cannot and they are not appropriate to make decision on behalf of the user the acceptance of the content. Only the user can decide if the content generated is appropriate, logical and correct. Usually, user has to work together with the *content generation assistant* to complete a content generation task.

Thus, it is better called the Midjourney an image generation assistant. ChatGPT is a text content generation assistant. Google Translate is a translation assistant. Google Search has a search sentence recommender. Cortona is a voice-to-text generation assistant. They play a role to support the user to generate content.

Question: Should an auto-driving system be designed as a system to assist a driver in controlling a vehicle only?

2.1 Image generation

Consider that image generation is an intelligent application provided by Midjourney. To use Midjourney to generate an artificial image, one needs to enter a text description about the image. Midjourney will need to *understand the meaning of the text description* and then *generate the values for the input variables* to the generative model for the *image generation*. Thus, the Midjourney would have at least to use two AI tools, one is for understanding the text description and the other is the generative model, for image generation in accordance with the user text description.

2.2 Text generation

Consider that text generation is an intelligent application provided by OpenAI ChatGPT and Google Bard, in which can now support multiple languages. To generate Chinese text in response to a Chinese prompt, the application systems would need to use at least two AI tools. One is used for Chinese-English translation and the other is used for English text generation, i.e. the large language model (LLM).

3 Information Infrastructure

To ensure that an intelligent service can be accessible by any user. Connections among the users' devices and the cloud platforms providing intelligent services must be ensured. Moreover, it should be possible to let the user to store the result generated by an intelligent service for record. It should also be possible to let a user to share the result to other users. To accomplish all these goals, an information infrastructure is inevitable.

Today, various technologies have been developed to support this information infrastructure. Below lists a number of them.

- Edge device (user-side): Notebook computer, cell phone, pad and many wearable device.
- Server-side: Computers (resp. servers) and specialized designed computers.
- Software: OS, network OS, database management system, file management system and others.
- Cloud platforms which provide a large amount of computing machines and memory space for users.
- Internet to support data communication among the computers and edge devices connected in this world-wide network.
- Telecom networks to support data and voice communication among the cell phones, the telephones and the computing machines connected in a country-wide (resp. city-wide) network.
- Satellite networks to support telecommunication and Internet services.

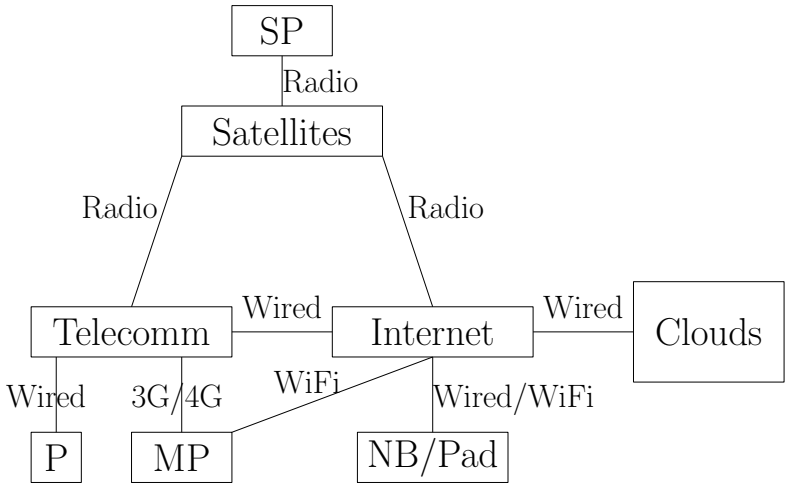

Figure 1: Architecture behind the information infrastructure which includes three major networks, namely (i) the Internet, (ii) the telecommunication network and (iii) the satellite network. The clouds are wired connected to the Internet. The telecommunication network is wired connected to the Internet. The connection between a telecommunication network (resp. Internet) and a satellite network is based on radio signal communication. In principle, a user can then connect to the information infrastructure by using his/her device, i.e. an edge device. The device can be a satellite phone (SP), a landline phone (P), a mobile phone (MP), a notebook computer (NB) and a Pad.

One point should be added. Connection between two devices can be accomplished by two types of technology, namely wired communication technology and wireless communication technology. For wired communication, a data is sent to another device in the form of electrical signal[1]. For wireless communication, a data is sent to another device in the form of radio signal.

3.1 User-Cloud connection

To ensure that a user is able to have such intelligent services accessible, *information infrastructure* is inevitable. Today, *information infrastructure* is a huge infrastructure which includes (i) various *cloud* platforms, (ii) the Internet, (iii) telecommunication networks and (iv) satellite networks. Figure 1 shows an architecture behind the *information infrastructure*.

3.1.1 Cloud platform

A cloud platform is in essence a network of computers (servers) which are wired connected to each other to manifest a giant computer with huge computation power and memory space. In this regard, various programs to be running to support the AI tools can be executed in the cloud platforms. Take the voice assistant appeared in 2011 as an example, voice-to-text generation was a computationally intensive task and the computation power of a mobile phone at that time was unable to handle such task. In the end, the computational task could only be conducted on the cloud platform. Without connection to the cloud platform via Internet, a mobile phone user was unable to get service from any voice assistant. Figure 2 shows an illustration on the connections among the cloud platforms and the Internet.

[1]It should be noted that one more wired communication technology is to send data to another device in the form of light via optical fibre. As this technology is usually applied in the data communication between routers, I do not emphasize this point in the main text owing not to complicate the introduction.

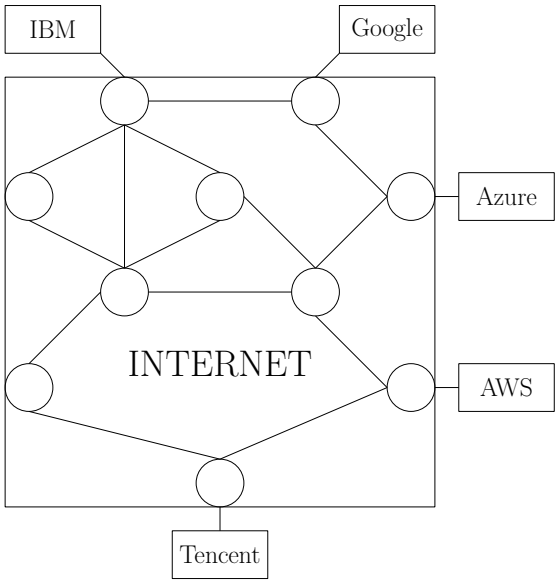

Figure 2: Cloud platforms are connected to the Internet to support their intelligent services.

3.1.2 Internet

The Internet is giant (worldwide) network of computers which are wired/wireless connected. Each computer is assigned with a unique ID called IP address. The main purpose of the Internet is simply to ensure that a data sent from a source computer (source IP address) can be received without error to the destination computer (destination IP address). In simple words, the main purpose of the Internet is to ensure *data communication* without any error.

3.1.3 Telecom network

Telecommunication network is a network laid by a telecommunication firm, like ChungHwa Telecom, to support **data communication service** *and* **voice communication service**. This network includes both a wired network connecting the landline telephone devices and a wireless network connecting the mobile phones. With the telecommunication technology advancement, all telecommunication networks are able to connect to the Internet for *data communication*. With the *data service* supported and the advancement on the mobile phones, a mobile phone user is now able to connect to the Internet and hence the *cloud platform*. Eventually, a mobile phone user is able to get the *intelligent services* provided by any *cloud platform*.

4 Evolution of Telecommunication Network

In early 19th century, telegraph network was invented to support data communication. In the late 19th century, telephone network was invented to support voice communication. To support telegraph and telephone services across multiple cities on the same continent, a large wired network was built. Owing to the advancement of the communication technology, the medium of signal transmission have been extended from wired media to wireless media.

- Wired: Copper cable, submarine cable, optical fibre, submarine optical fibre cable.
- Wireless: Short range radio signal on the earth, long range radio signal on the earth, long range radio signal among earth stations and satellites, ultra long range signal among earth

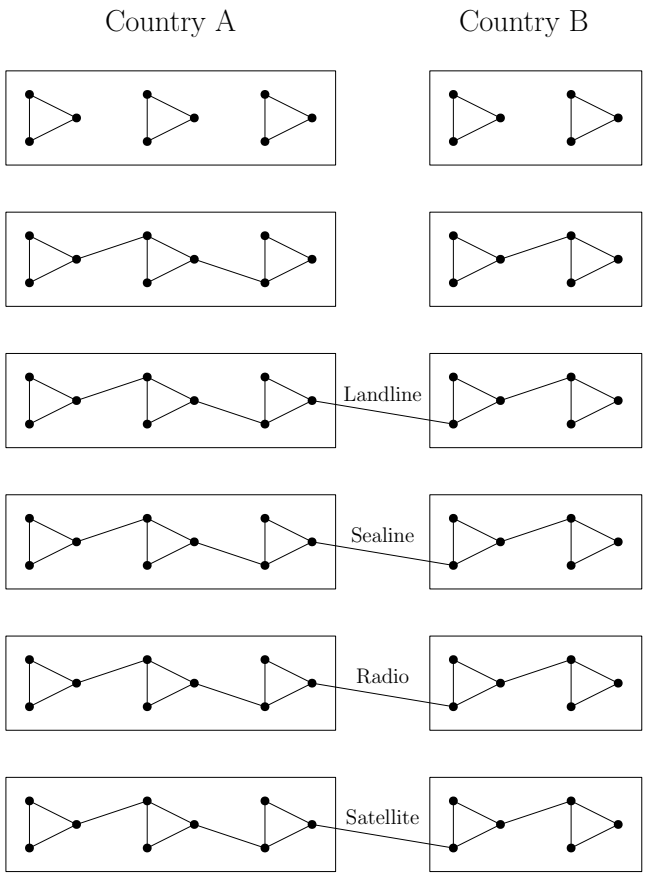

Figure 3: The key ideas behind the evolution of telecommunication networks. From top to bottom, (i) isolated city-wise (telegraph and telephone) network; (ii) city-wise networks are wired connected to form a country-wise network; (iii) country-wise networks are wired connected by land-lines; (iv) country-wise networks are wired connected by sea-lines; (v) country-wise networks are wireless connected by radio signals transmitted by base stations on earth; (vi) country-wise networks are wireless connected by radio signals transmitted by base stations on earth via satellites. In the end, the world-wide telecommunication network is formed.

stations and satellites in the outerspace.

Figure 3 shows the key ideas behind the evolution of telecommunication networks.

4.1 Wired telecom network

Clearly, telegraph and telephone services became demanding for cities across continents. With the technological advancement on the communication cables, submarine cables were laid in Europe and the Atlantic Ocean. At the turn of 20th century, major cities in the world had at least one post office which was able to provide telegraph (i.e. data communication) service. Many city-wise wired telephone networks were also built in many cities.

4.2 In need of radio communication

Figure 4 shows the key ideas behind the evolution of telecommunication based on radio signal. For two far-distance apart cities, say A and B, their base stations are unable to communicate each other via radio signal. To resolve this problem, the first method is to relay the radio signal via the

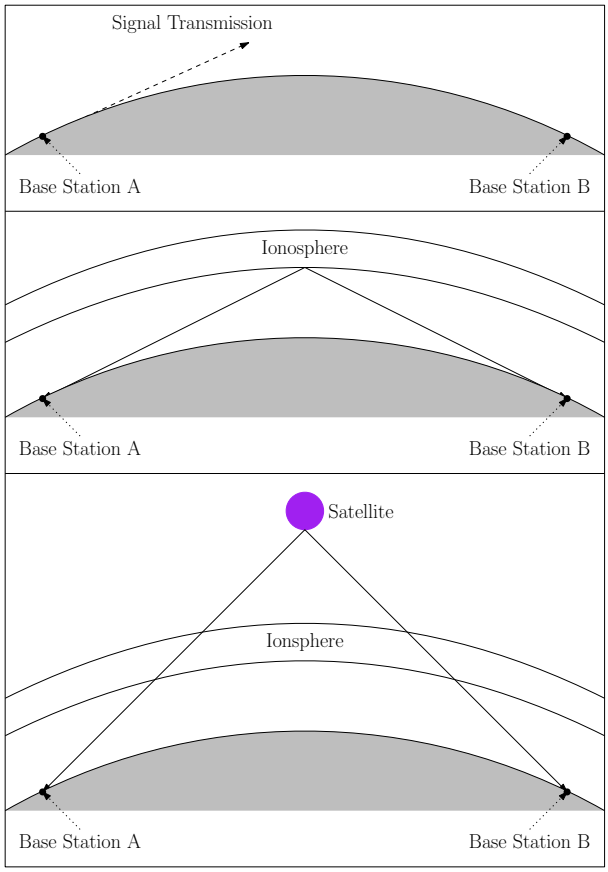

Figure 4: The key ideas behind the evolution of telecommunication based on radio signal. From top to bottom, (i) radio signal transmission is not possible for two base station which are far-distance apart, (ii) radio signal transmitted from one station is relayed by the ionosphere to another station, (iii) radio signal transmitted from one station is relayed by a satellite to another station.

ionosphere. The ionosphere acts as a mirror up on the sky reflecting radio signal from one station to another. It works for the base stations which are not very far apart. For two base stations which are very far apart, like a base station in Iceland and a base station in New Zealand, radio signal relaying by the ionosphere is poor. Thus, the second method is to relay the radio signal via a satellite.

4.3 Wireless telecom network

In the early 20th century, due to the advancement in radio communication, wireless communication was made possible. Thus, radio (i.e. wireless) communication technology was applied in (world-wide) telegraph network (resp. telephone network). It is evidenced that the communication technology to support voice communication is always more advanced than the technology to support data communication. Telegraph networks and telephone networks eventually merged. One single telecommunication network was able to support both *data* and *voice* communications.

Below lists the down-link (DL) and up-link (UL) speeds, in terms of bit-per-second (bps), of the generations from 2G to 5G. Remind that one byte is eight bits. For an image file its size is 1MB, its total number of bits is 8M bits. Thus, downloading a 1MB image file based on 3G technology requires about 0.2 second. If the image file is downloaded by a 5G network, the time is less than 0.003 second.

Generation	DL (bps)	UL (bps)
2G	85.6K	21.4K
	64.2K	42.8K
	236.8K	59.2K
	177.6K	118.4K
3G	42M	11M
4G	150M	50M
5G (Huawei)	3.6G	250M
5G-A	10G	1G

4.4 Satellite network

In 1959, the first satellite *Sputnik* for telecommunication was orbited around the earth by the Soviet Union. It marks a milestone on satellite communication, as the distance for a *data* or *voice* to be transmitted from one place on the earth to another was much longer.

Since after 1959, a large number of satellites have been delivered to the outerspace orbiting around the earth. Nowadays, some satellite networks are used for global position, so-called global positioning system (GPS). Some satellite networks are used for weather observatory. Some satellite networks are used for extending the services of a telecom network. Some satellite networks are used for extending the *data communication* service of the Internet.

4.5 Generations of mobile phone network

Mobile phone technology has been undergone a number of generations of advancement. In the 1980s, the first generation of mobile phone technology was developed. The second generation (2G) was released and widely used in the 1990s. Then, 3G and 4G wireless telecom technologies were introduced in subsequent decades.

With the advance of wireless telecom technology, the number of data and the size of a data to be communicated are large. Thus, the technologies developed to support telecom service are shifted to support data communication. No matter to communicate a text, a voice or a video over a telecom network, the key technology is first to convert the text, the voice or the video to data. On the receiver side, the data is then converted back to the text, the voice or the video.

4.6 Future telecommunication network

Since 1990s, there have been a number of on-going researches on Interplanetary data communication[2]. Imagine that you would like to send an email to a friend who is located in Mars. How long you will expect that the email will arrive your friend's email account?

In a NASA space exploration project, a data to be sent from an outerspace vehicle to an earth station is based on two approach. First, it is based on line-of-sight direct communication. It works if a vehicle located on the earth-facing of the moon can send radio signal directly to an earth station. Second, it requires a satellite orbiting above moon for relaying the signal. It works if a vehicle located on the opposite side cannot send radio signal directly to the earth. Then, at least one satellite is needed to orbit the moon to relay the radio signal transmission from the moon to the earth. In either cases, the energy consumption in supporting such communication is large and thus either approach is expensive.

Owing to trade-off the delay of a data transmission, the number of satellites orbiting in the outerspace and hence the cost of transmission, one design is based on the idea of delay-tolerant. The outerspace satellite network supporting such service is called delay-tolerant network (DTN), see Figure 5.

[2] `https://en.wikipedia.org/wiki/Interplanetary_Internet`.

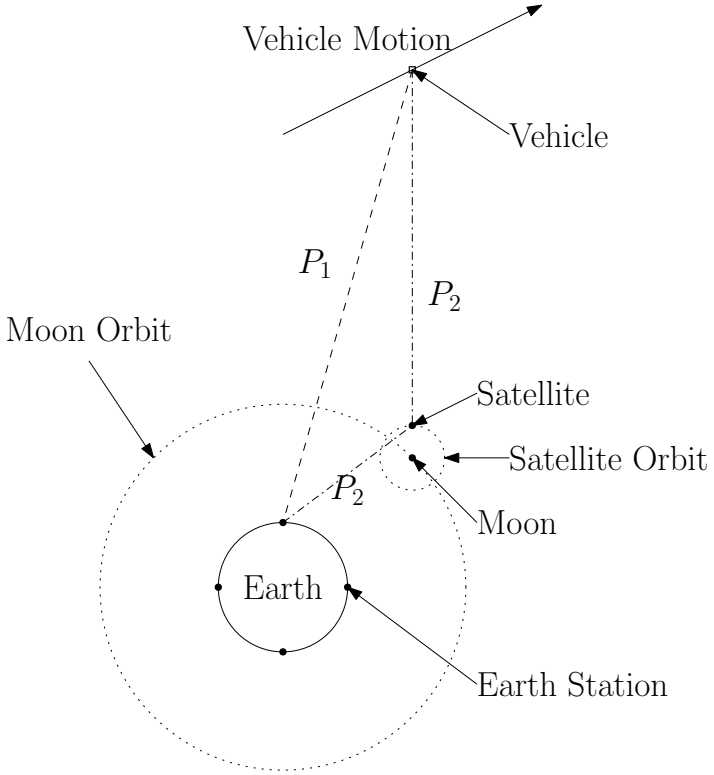

Figure 5: Communication between an outerspace vehicle and an earth station. If the communication is based on direct line-of-sight (P_1), the power consumption is higher. If the communication is based on P_2 in which a moon satellite replaying the signal to an earth station, the power consumption is lower.