中国科学院大学
University of Chinese Academy of Sciences

CS101

网络思维-3

网络规律，职业素养

zxu@ict.ac.cn

提纲

- 网络思维概述、名词术语
- 网页编程
- 连通性与协议栈

- 网络规律与职业素养
  - 互联网简史：发展与社会影响
  - 网络规律
    - 延时与带宽的霍克尼公式
    - 网络效应：梅特卡夫定律、里德定律、病毒市场现象
  - 职业素养
    - 安全，隐私，职业规范

课件中包含教科书未包括的素材引用，特此致谢

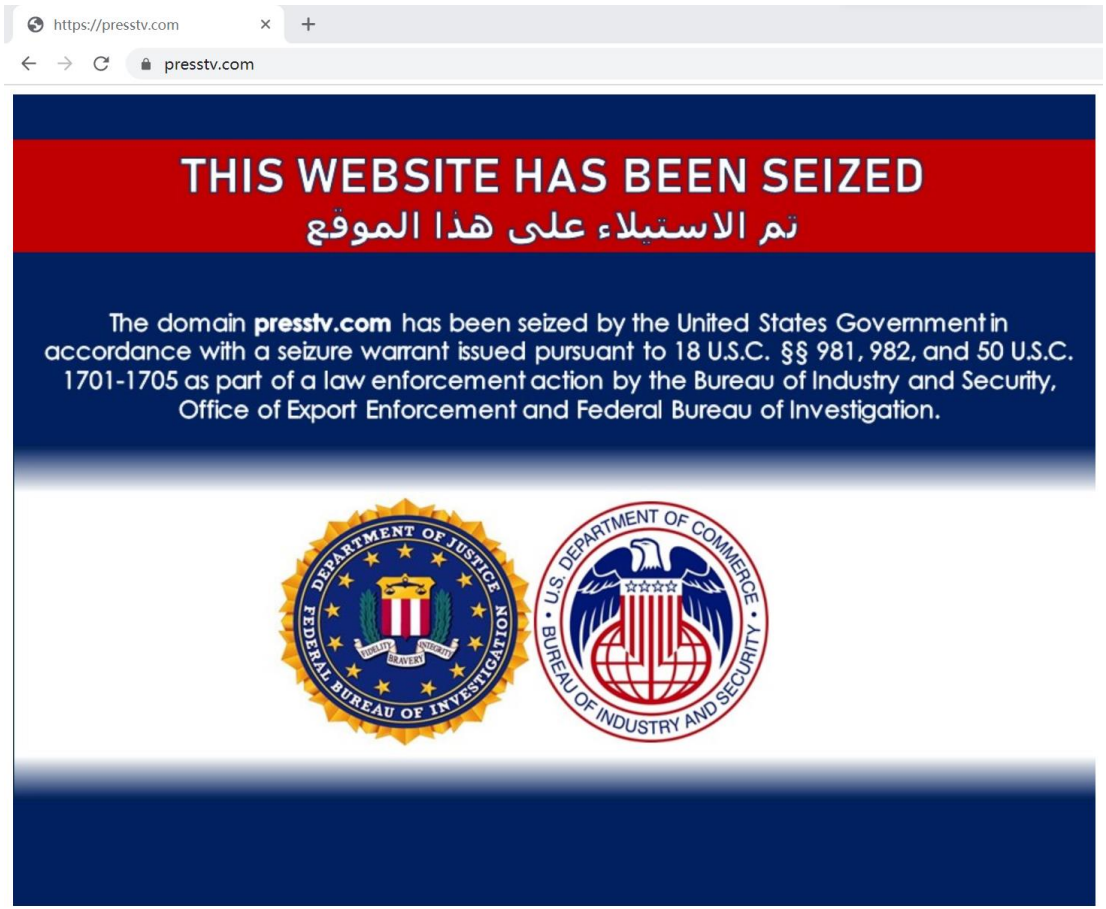
1. 计算机网络简史

- 不断的进步与演化
  - 连通性（什么连在网上），协议栈（传递什么）

Start Time	Milestone 里程碑事件	Main Functions 主要功能
1800's	Telecommunication networks	Telephony, telegraph 电信网：电话电报
1963	J. C. R. Licklider proposed the concept of Intergalactic Computer Network	A general idea of computer networks 计算机网络思想
1969	First message sent on ARPANET 阿帕网 (50 Kbps = 50 Kilo bits per second)	Message passing, packet switching, interface HW 信息传递 分组交换
1974	TCP/IP	Internetworking (Internet) 因特网 with telnet, ftp, email applications
1989	World Wide Web 万维网	More applications enabled by global-scale hypertexting 超文本、超链接
2000	Network science, 网络科学，网格计算 grid computing, cloud computing 云计算	Various networks as the object of scientific inquiry, utility computing
2007	Apple iPhone 苹果公司智能手机	Mobile Internet 移动互联网
2008	Bitcoin 比特币	Blockchain, network of trust 区块链
2021	美国封禁伊朗网站事件	断网威胁

伏羲智库“域名防篡改监测服务”复盘美国封禁伊朗网站事件（2021-06-24发布结果）

- 2021年6月22日，美国政府以“违反制裁”为由关闭了包括伊朗英语新闻电视台（PressTV.com）等在内的36家与伊朗相关的新闻网站



伏羲智库“域名防篡改监测服务”复盘美国封禁伊朗网站事件（2021-06-24发布结果）

- 伏羲智库“域名防篡改监测服务”
  - 基于部署在全球超过300个服务器上的DNS数据采集网络，
    收集由递归DNS服务器
    与权威DNS服务器间实时查询响应报文，
    并进行实时分析
    - 每天采集超过 2TB 的DNS解析数据
    - 每秒钟处理超过20万次DNS记录
    - 达到1Gb/秒的实时数据流采集处理
    - 已积累过去十年间超过1000亿条历史DNS解析记录
  - 可实现对网站域名、IP地址的解析服务的完整性性、安全度进行实时分析，对域名劫持、DNS篡改、DNS投毒、DNS污染等针对企业及机构网站域名、IP地址的威胁进行实时监控

伏羲智库“域名防篡改监测服务”复盘美国封禁伊朗网站事件（2021-06-24发布结果）

- 媒体报道中涉及的presstv.com、alalamtv.net、almasirah.net 三个网站域名，在6月22日之前都分别指向各自域名解析服务器，并稳定运行
- 在UTC时间2021年6月22日同一时间段，三个域名的域名解析服务器，被修改至四个相同的域名解析服务器
  - ns-388.awsdns-48.com
  - ns-977.awsdns-58.net
  - ns-1088.awsdns-08.org
  - ns-1900.awsdns-45.co.uk
- 经查该四个域名解析服务器都是美国公司亚马逊aws云提供的域名解析服务器

伏羲智库“域名防篡改监测服务”复盘美国封禁伊朗网站事件（2021-06-24发布结果）

- 以被封禁的presstv.com 为例
- 在6月22日前，presstv.com网站一直解析到IP地址77.66.40.12所在服务器
  - 通过IP地址位置数据库查询，该服务器部署在位于丹麦首都大区霍耶－措斯楚普自治市的数据中心
- 6月22日起， presstv.com 域名解析A记录开始出现频繁变化，指向13.249.79.2、13.249.118.12、 52.85.242.35、99.84.189.3 等IP地址，都属于美国亚马逊机房的IP地址
- Verisign是顶级域.com下所有二级域（包括presstv.com）的域名注册管理机构
  - 即，美国公司Versign是presstv.com 域名数据的管理机构

伏羲智库“域名防篡改监测服务”复盘美国封禁伊朗网站事件（2021-06-24发布结果）

- 综上初步分析
  - 技术层面美国司法部通过顶级域名注册管理机构，将所涉及网站的域名解析A记录和NS记录强制指向美国政府指定的亚马逊域名解析服务器，从而实现对指定网站进行关停处理

  - A记录：用于指定域名对应的IP地址
    - 例如，presstv.com的A记录用于指定网站域名对应的提供网站服务的主机IP地址

  - NS记录：用于指定提供该域名的解析服务的域名服务器名称
    - 例如，presstv.com的NS记录用于指定解析presstv.com域名的服务器名称

Internet history is a history of social impact
互联网产生越来越大的社会影响

- With technology advances comes increasing social impact
  - The size of the global Internet has grown exponentially
    - The trend is likely to continue till 2050
  - Most of nodes are hosts (edge nodes), not networking devices

Evolution of Internet 互联网规模增长：历史与展望

Time 时间	# Nodes 节点数	Exemplar Techniques 标志性技术
1960s	A few 数个	Packet Switching Network 分组交换网
1970s	Thousands 数千个	TCP/IP, Ethernet 因特网、以太网
1980s	100 thousands 数十万	Client-Server Computing 客户-服务器计算
1990s	Millions 数百万	World Wide Web 万维网
2000s	100 Millions 数亿	Cloud Computing 云计算
2010s	Billions 数十亿	Smartphones, Mobile Internet 移动互联网
2020-2050	Trillions 数万亿	Internet of Human-Cyber-Physical Systems 人机物融合的智能互联

2. 网络规律

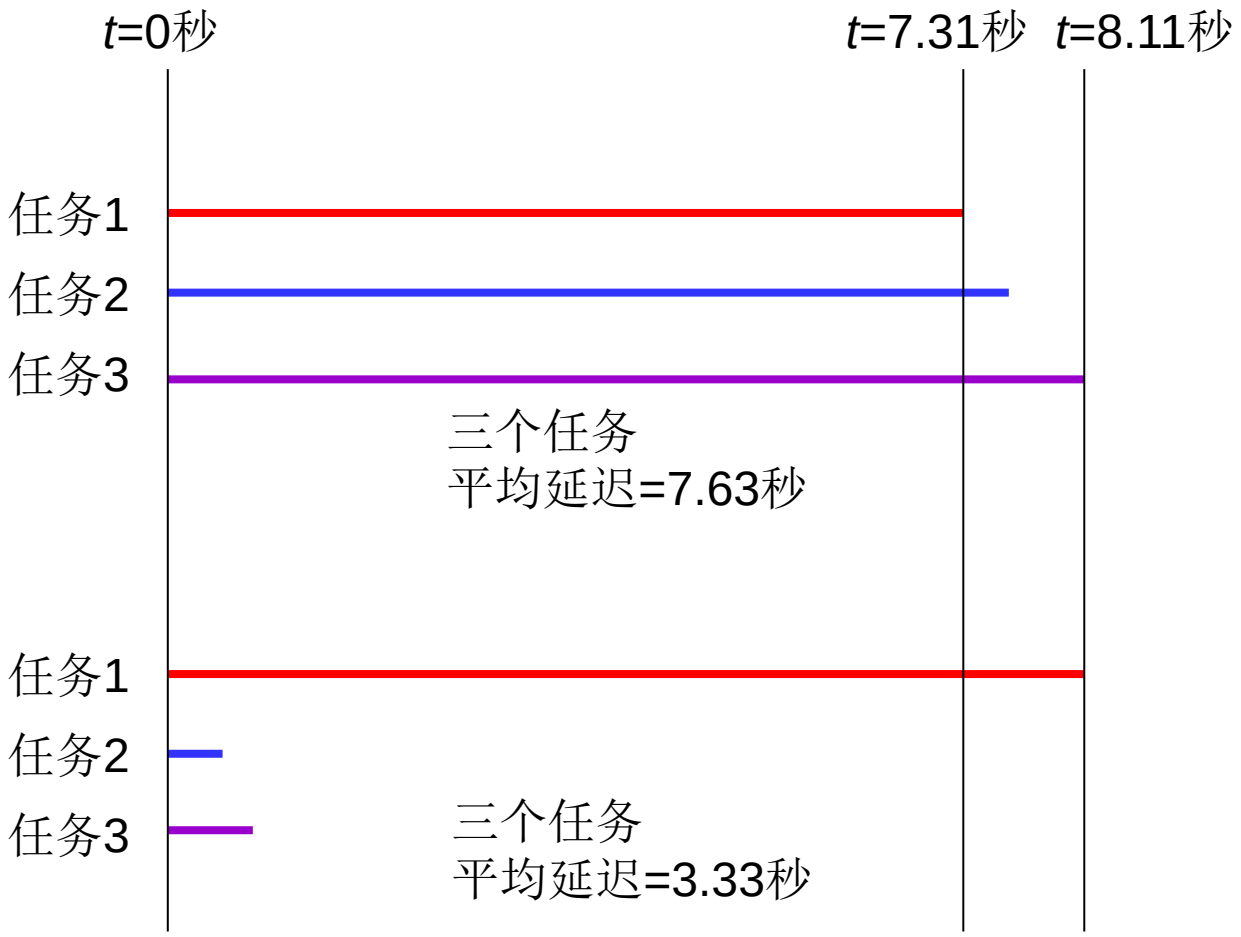
- 网络规律是像摩尔定律一样的观察，不是物理学定律
  - 网络的延时与带宽的霍克尼公式
  - 网络效应
    - Metcalfe’s law 梅特卡夫定律
    - Reed’s law 里德定律
    - The Viral marketing phenomenon 病毒市场现象

假设三个下载任务使用10 Mbps带宽资源

平均延迟更小

任务1没有阻塞其他任务

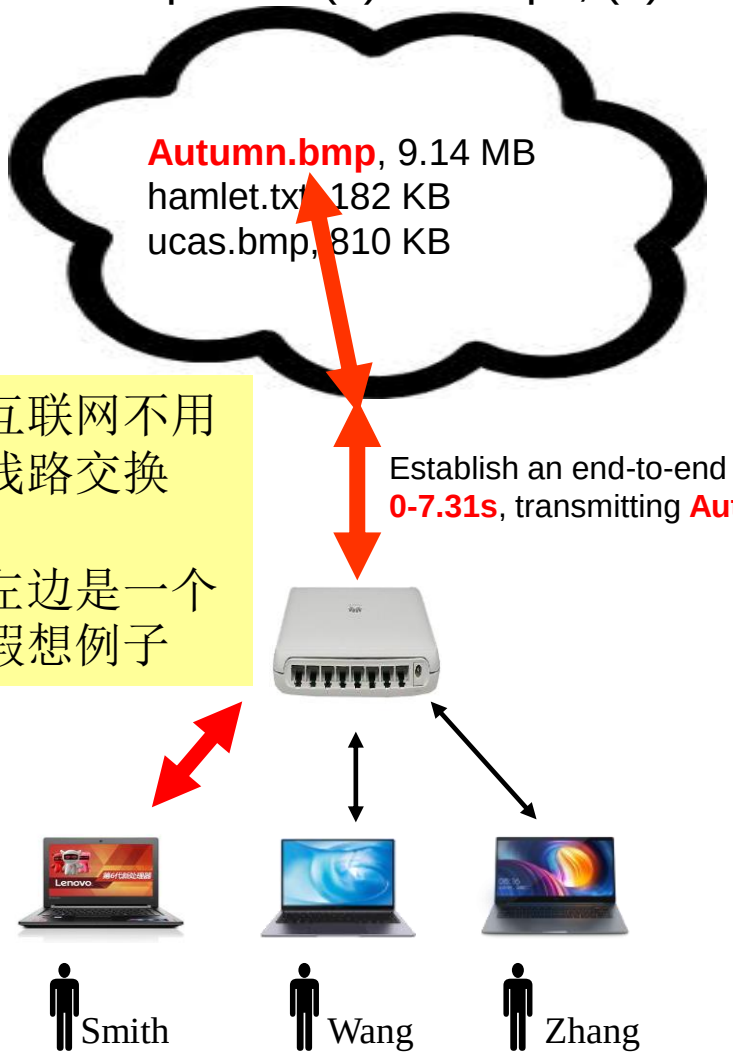
Assumptions for both systems:
(1) 10 Mbps; (2) all three tasks start at 0;
(3) **ignore all overheads**

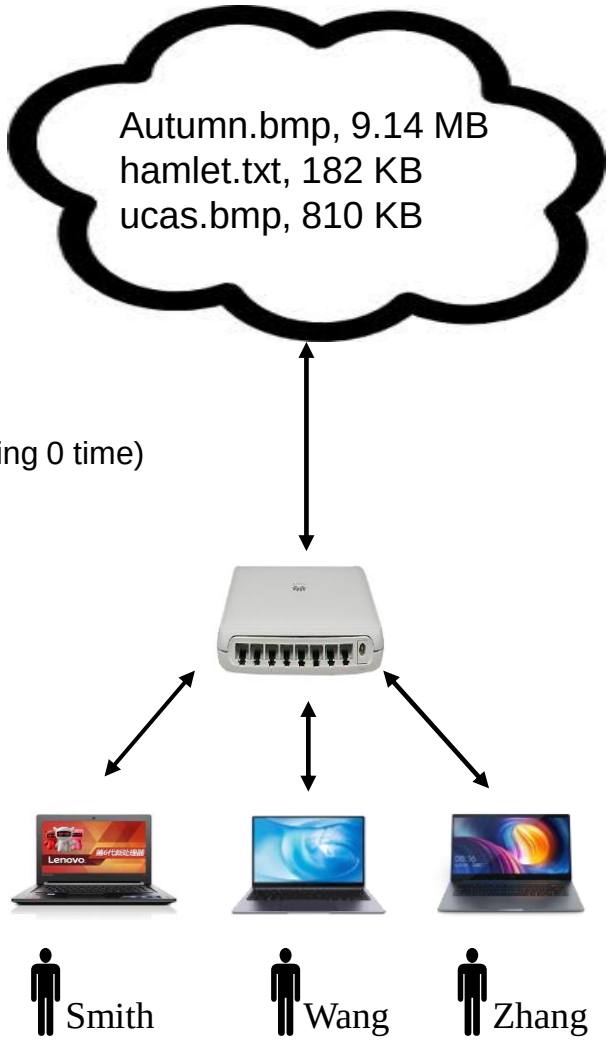


复习：

Circuit switch vs. packet switch

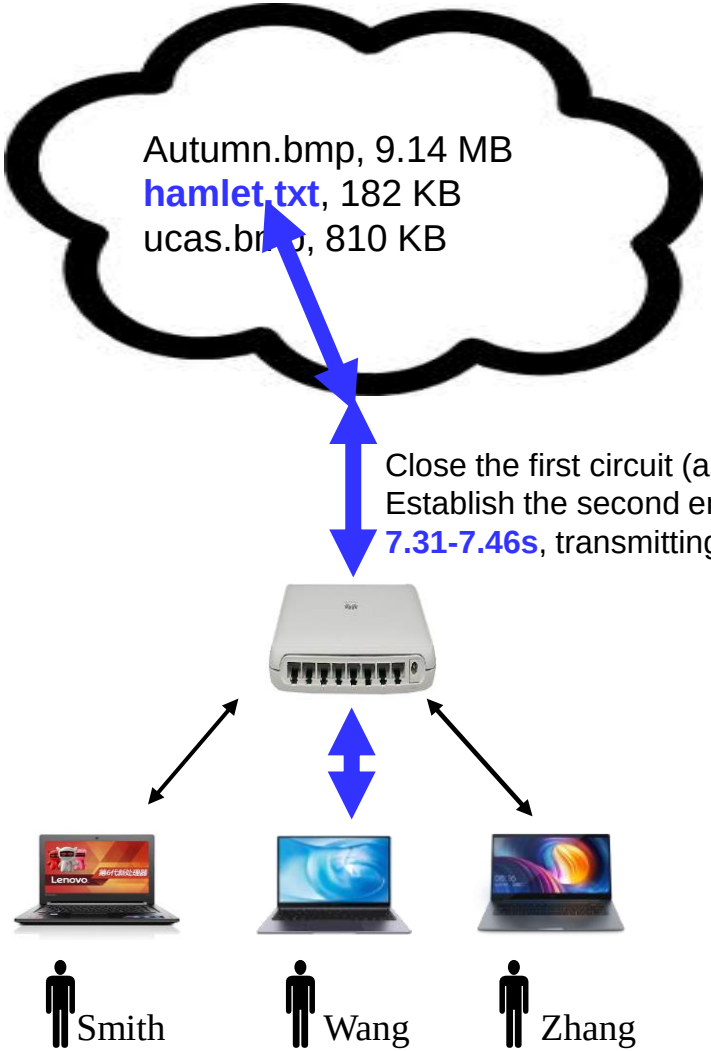
Assumptions: (1) 10 Mbps; (2) all three tasks start at 0; (3) ignore all overheads

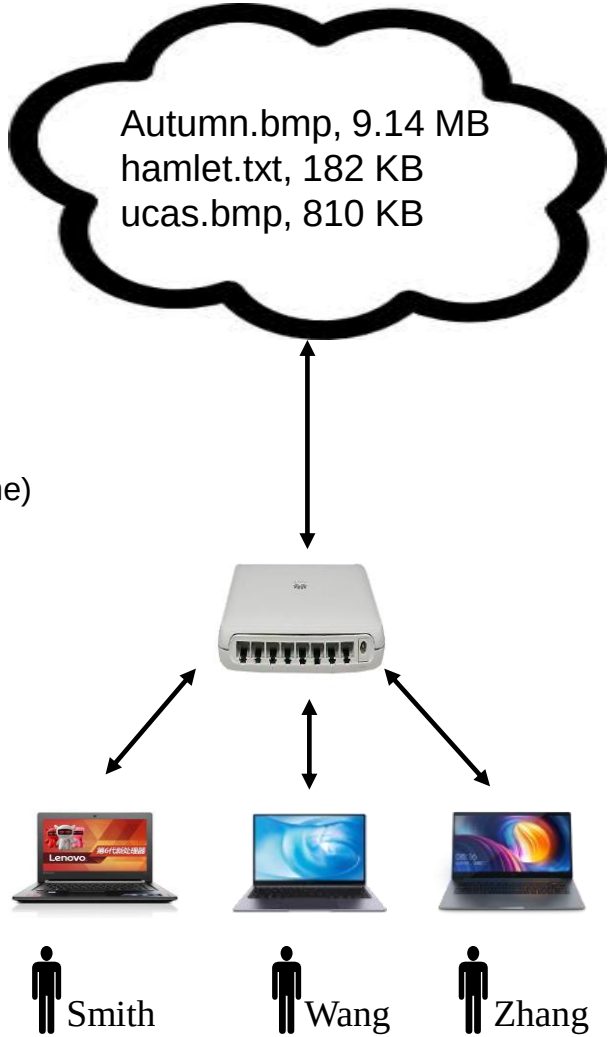




Circuit switch vs. packet switch

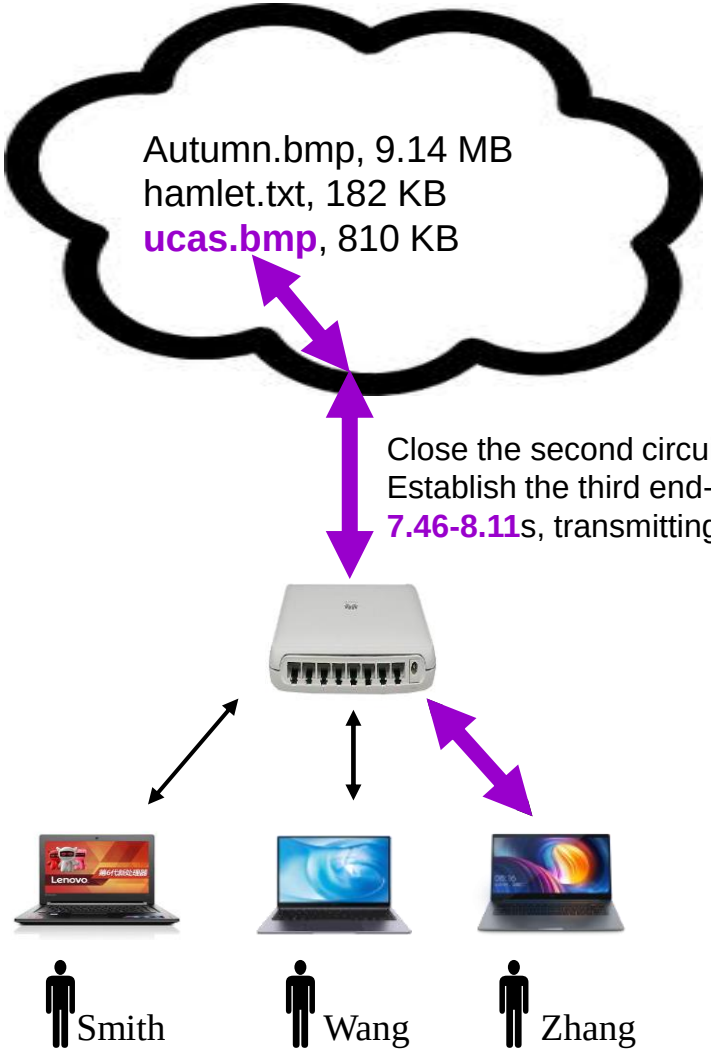
Assumptions for both systems:
(1) 10 Mbps; (2) all three tasks start at 0; (3) ignore all overheads

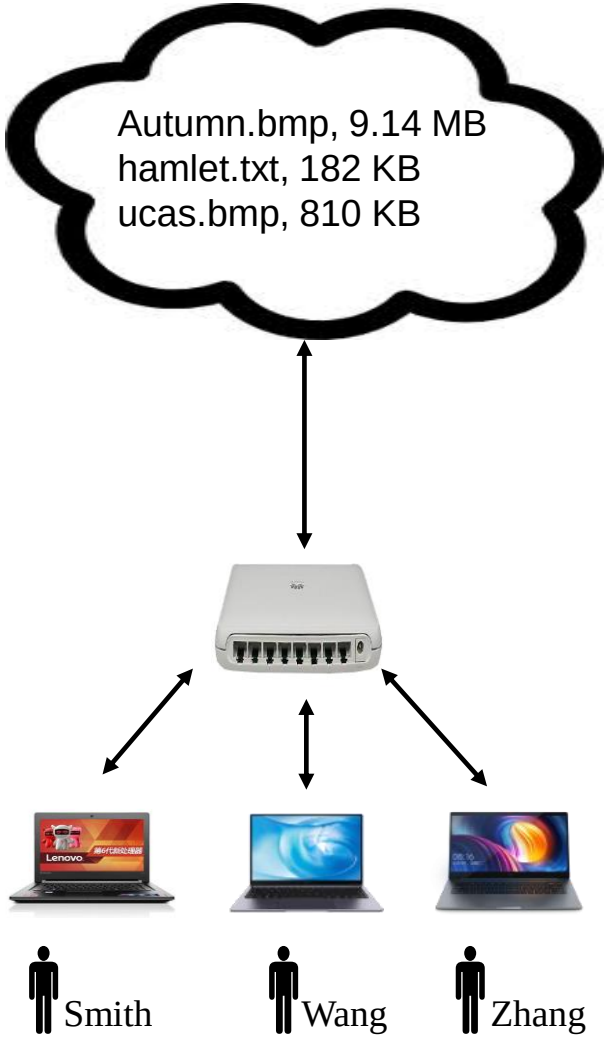




Circuit switch vs. packet switch

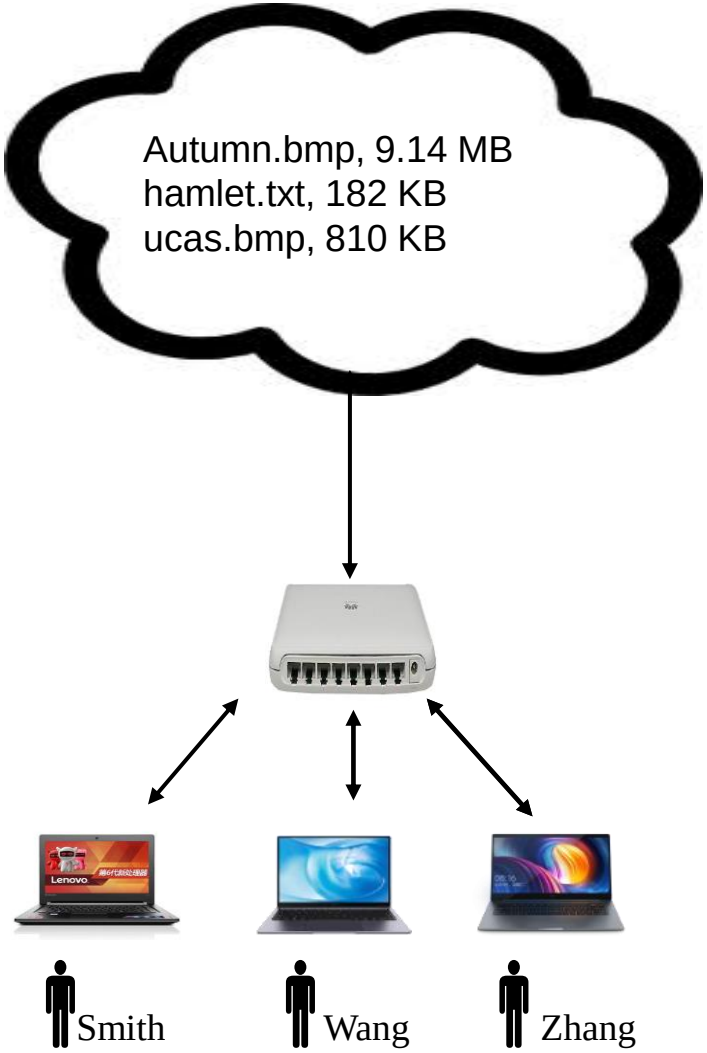
Assumptions for both systems:
(1) 10 Mbps; (2) all three tasks start at 0; (3) ignore all overheads



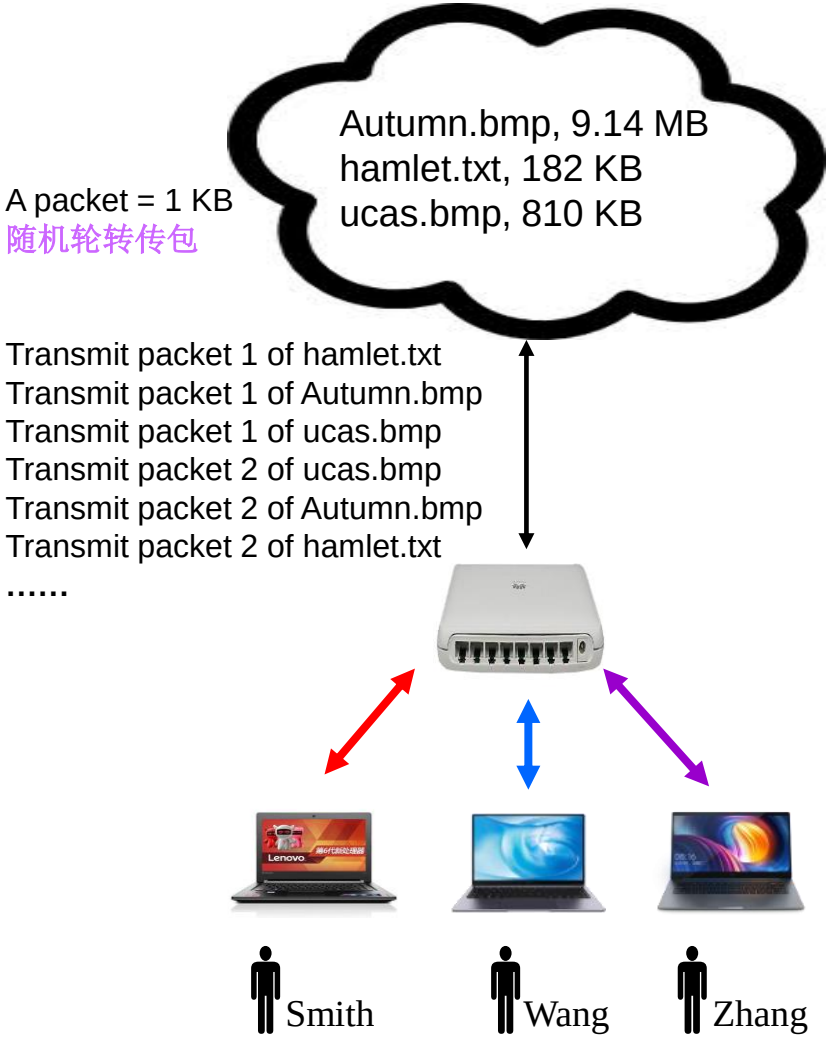


Circuit switch vs. packet switch

Assumptions for both systems:
(1) 10 Mbps; (2) all three tasks start at 0; (3) ignore all overheads

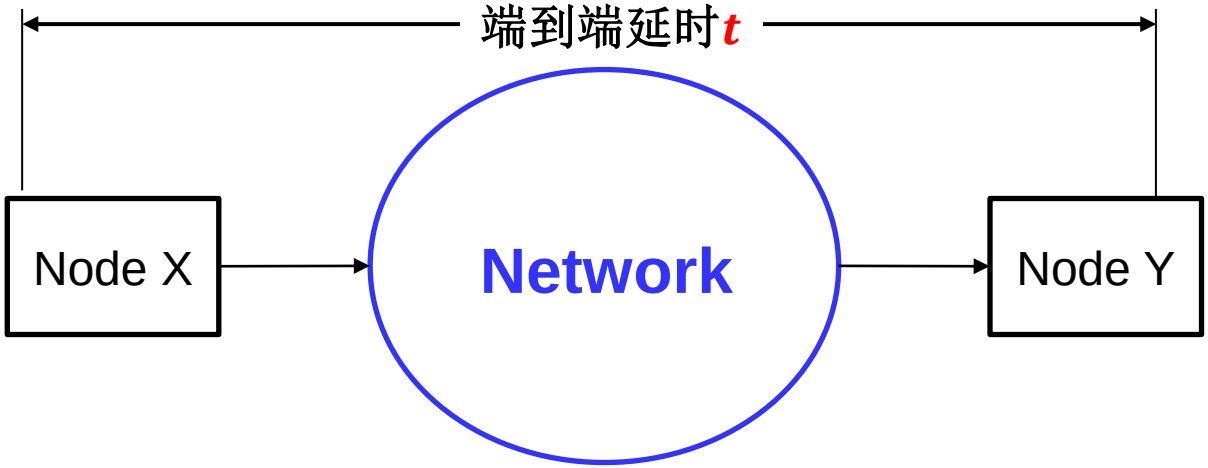


A packet = 1 KB
随机轮转传包



2.1 Latency and bandwidth 延时与带宽

- 最简单的通信：节点X传送一条m字节的消息到节点Y
  - 延时：the total time t to transmit the message
  - 带宽：number of bits transmitted per second (bps)
- 霍克尼近似公式 Hockney's formula: 延时$t = t_0 + m/r_\infty$
  - Extreme values 极端值
    - 最小延时 Minimal latency: $m = 0$时的延时$t = t_0$ 空消息延时
    - 最大带宽 maximal bandwidth: $m \rightarrow \infty$时的带宽r_∞ 大消息带宽
  - User experienced values 用户体验值
    - User experienced latency: t; User experienced bandwidth m/t



Example: network hero experiments data
展示最大带宽的通信实验

- 1-minute quiz
  - Q1: How much time is needed to transmit a movie file of 1GB over the hero network of 2013?
  - Q2: How much time is needed to transmit a text file of 1KB over the hero network of 2013?

Time of Experiment	Maximal Bandwidth Achieved r_∞	Time t to Transmit 1 GB		Time t to Transmit 1 KB	
		$t_0 = 1\ \mu s$	$t_0 = 1\ ms$	$t_0 = 1\ \mu s$	$t_0 = 1\ ms$
1975	4.50E+07 bps, or 45 Mbps				
1984	1.00E+09 bps, or 1 Gbps				
1993	1.53E+11 bps, or 153 Gbps				
2002	1.00E+13 bps, or 10 Tbps				
2013	8.18E+14 bps, or 818 Tbps	?	?	?	?

Example: network hero experiments data

- 1-minute quiz
  - Q1: How much time is needed to transmit a movie file of 1GB over the hero network of 2013?
    - A1: 8 Gb / 818 Tbps = 9.78E-06 = 9.78 **μs** $t = m/r_\infty$
  - Q2: How much time is needed to transmit a text file of 1KB over the hero network of 2013?
    - A2: 8 Kb / 818 Tbps = 9.78E-12 = 9.78 **ps** $t = m/r_\infty$

Time of Experiment	Maximal Bandwidth Achieved r_∞	Time t to Transmit 1 GB		Time t to Transmit 1 KB	
		$t_0 = 1$ μs	$t_0 = 1$ ms	$t_0 = 1$ μs	$t_0 = 1$ ms
2013	8.18E+14 bps, or 818 Tbps	9.78 **μs**	9.78 **μs**	9.78 **ps**	9.78 **ps**

Example: network hero experiments data

- 1-minute quiz
  - Q1: How much time is needed to transmit a movie file of 1GB over the hero network of 2013?
    - A1: 8 Gb / 818 Tbps = 9.78E-06 = 9.78 **μs** $t = m/r_\infty$
  - Q2: How much time is needed to transmit a text file of 1KB over the hero network of 2013?
    - A2: 8 Kb / 818 Tbps = 9.78E-12 = 9.78 **ps** $t = m/r_\infty$
  - What is wrong with these answers?
    - Did not consider t_0, the startup overhead
    - The time to transmit a 0-byte message is not 0 second!

空消息传输时间 ≠ 0！

Time of Experiment	Maximal Bandwidth Achieved r_∞	Time t to Transmit 1 GB		Time t to Transmit 1 KB	
		$t_0 = 1$ μs	$t_0 = 1$ ms	$t_0 = 1$ μs	$t_0 = 1$ ms
2013	8.18E+14 bps, or 818 Tbps	9.78 μs			9.78 ps

Example: network hero experiments data
更接近正确的答案：使用$t = t_0 + m/r_\infty$

- 1-minute quiz: correct answers
  - Q1: How much time is needed to transmit a movie file of 1GB over the hero network of 2013?
    - A1: If $t_0 = 1$ **μs**, t = 1E-06 + 8 Gb / 818 Tbps = 10.78E-06 = 11 **μs**
      If $t_0 = 1$ **ms**, t = 1E-03 + 8 Gb / 818 Tbps = 1E-03 = 1 **ms**
  - Q2: How much time is needed to transmit a text file of 1KB over the hero network of 2013?
    - A2: If $t_0 = 1$ **μs**, t = 1E-06 + 8 Kb / 818 Tbps = 1E-06 = 1 **μs**
      If $t_0 = 1$ **ms**, t = 1E-03 + 8 Kb / 818 Tbps = 1E-03 = 1 **ms**

Time of Experiment	Maximal Bandwidth Achieved r_∞	Time t to Transmit 1 GB		Time t to Transmit 1 KB	
		$t_0 = 1$ μs	$t_0 = 1$ ms	$t_0 = 1$ μs	$t_0 = 1$ ms
2013	8.18E+14 bps, or 818 Tbps	11 μs	1 ms	1 μs	1 ms

Example: network hero experiments data
用户体验到的带宽是多少？

- 1-minute quiz: correct answers 使用$t = t_0 + m/r_\infty$
  - Q1: transmit 1GB over the hero network of 2013?
    - A1: If $t_0 = 1$ **µs**, t = 1E-06 + 8 Gb / 818 Tbps = 10.78E-06 = 11 **µs**
      If $t_0 = 1$ **ms**, t = 1E-03 + 8 Gb / 818 Tbps = 1E-03 = 1 **ms**
  - Q2: transmit 1KB over the hero network of 2013?
    - A2: If $t_0 = 1$ **µs**, t = 1E-06 + 8 Kb / 818 Tbps = 1E-06 = 1 **µs**
      If $t_0 = 1$ **ms**, t = 1E-03 + 8 Kb / 818 Tbps = 1E-03 = 1 **ms**
    - **短消息带宽低：*m / t* = 1KB / 1ms = 8 Mbps，远低于818 Tbps**

Time of Experiment	Maximal Bandwidth Achieved r_∞	Time t to Transmit 1 GB		Time t to Transmit 1 KB	
		$t_0 = 1$ µs	$t_0 = 1$ ms	$t_0 = 1$ µs	$t_0 = 1$ ms
2013	8.18E+14 bps, or 818 Tbps	11 µs	1 ms	1 µs	1 ms
	用户体验带宽 *m/t* =	742 Tbps	7.92 Tbps	741 Mbps	7.92 Mbps

Lessons learned regarding $t = t_0 + m/r_\infty$

- For short messages, the first term $\boldsymbol{t_0}$ often dominates
- For a long message, the second term $\boldsymbol{m/r_\infty}$ often dominates
- User experienced latency is $\boldsymbol{t}$, not $\boldsymbol{m/r_\infty}$
- User experienced bandwidth is $\boldsymbol{m/t}$, not $\boldsymbol{r_\infty}$
  - To transmit a 1-KB message over a network with $r_\infty = 1$ Gbps and $t_0 = 1\ ms$, the user experienced bandwidth is m/t = 7.9 Mbps, 125 times smaller than the maximal bandwidth $r_\infty = 1$ Gbps

Time of Experiment	Maximal Bandwidth Achieved r_∞	Time t to Transmit 1 GB		Time t to Transmit 1 KB	
		$t_0 = 1\ \mu s$	$t_0 = 1\ ms$	$t_0 = 1\ \mu s$	$t_0 = 1\ ms$
1975	4.50E+07 bps, or 45 Mbps	178 s	178 s	0.2 ms	1.2 ms
1984	1.00E+09 bps, or 1 Gbps	8 s	8 s	9 μs	1 ms
1993	1.53E+11 bps, or 153 Gbps	52 ms	53 ms	1.1 μs	1 ms
2002	1.00E+13 bps, or 10 Tbps	0.8 ms	1.8 ms	1 μs	1 ms
2013	8.18E+14 bps, or 818 Tbps	11 μs	1 ms	1 μs	1 ms

1-minute quiz

- Q: Paid for 1-Gbps and got only 5 Mbps. Why?
  从电信公司买了1-Gbps，只得到5 Mbps。为什么？
  - I subscribe to a fiber optical plan from a reputable ISP, which offers a 1-Gbps bandwidth connection to the Internet. However, I often only experience 5 Mbps bandwidth when accessing the Internet. Why this huge (up to 200 times) disparity?

1-minute quiz

- Q: Paid for 1-Gbps and got only 5-8 Mbps. Why?
  - I subscribe to a fiber optical plan from a reputable ISP, which offers a 1-Gbps bandwidth connection to the Internet. However, I often only experience 5 Mbps bandwidth when accessing the Internet. Why this huge (up to 200 times) disparity?
- A: The following are possible reasons
  - The 1-Gbps connection is only part of the full path from my laptop to the servers on the Internet. Some parts of the rest of the path are slower than 1-Gbps. 只有接入是1-Gbps
  - I am sharing the 1-Gbps connection with other students. 1-Gbps接入不是独占，而是与其他用户分享
  - I am accessing a lot of small files, resulting in transmissions of many short messages. Thus, the user experienced bandwidth is smaller than the maximal bandwidth 1-Gbps. 用户传输大量短消息，实际体验到的带宽远低于最大带宽1-Gbps

Compression 数据压缩

- Data compression: Technique to reduce file size
  - To save storage space and transmission time
- Lossless compression 无损压缩
  - Reduce file size without losing information
    - > gzip fib-10 (2011793 bytes)
    - To obtain a compressed file fib-10.gz (709090 bytes)
    - > gzip Autumn.bmp (9144630 bytes)
    - The compressed file is Autumn.bmp.gz (8224455 bytes)
  - Original file can be recovered from compressed file
    - > gzip -d Autumn.bmp.gz
- Lossy compression 有损压缩
  - Reduce file size while losing information
  - Original file cannot be recovered from compressed file

Lossy compression

- Original file

  > ls -l Autumn.png

  -rw-r--r-- 1 5971405 Autumn.png

5971405
5.97 MB

- Reduce size ~10 times

  > pngquant --quality=1 Autumn.png

  > ls -l Autumn-fs8.png

  -rw-r--r-- 1 597384 Autumn-fs8.png

597384
597 KB

- Reduce size ~64 times

  > pngquant --quality=0 Autumn.png

  > ls -l Autumn-fs8.png

  -rw-r--r-- 1 92506 Autumn-fs8.png

92506
92.5 KB

2.2 Network effect 网络效应

- Network laws

  n = number of nodes

  - **Metcalfe's law** ($V \propto n^2$) 梅特卡夫定律
    - Value *V* of a network of *n* nodes is proportional to n^2
  - **Reed's law** ($V \propto 2^n$) 里德定律
    - Value *V* of a network of *n* nodes can scale exponentially with *n*, because the network can form 2^n subgroups
- Viral marketing 病毒市场现象
  - Markets grow wide and fast, like biologic viruses
  - Why?
    - Connected and 0-cost
      - Zero purchasing price
      - Zero usage cost
      - Zero propagation cost

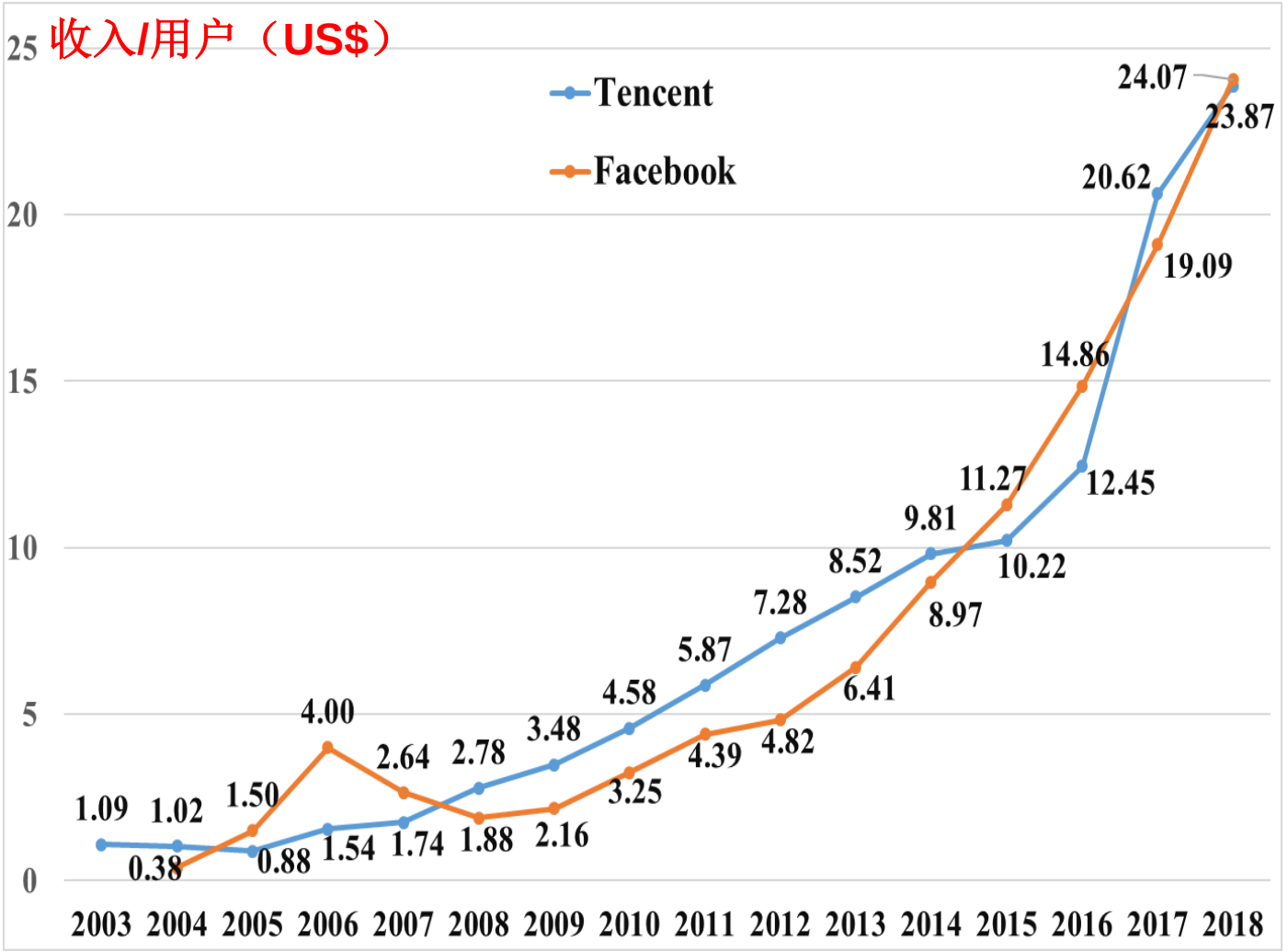
脸书网与腾讯网实例

- Facebook与Tencent的年报数据(2003-2020)，适配平方公式
  - 脸书网络的价值（年收入）=$10.13 \times 10^{-9} \times n^2$美元，误差=5.65%
  - 腾讯网络的价值（年收入）=$10.89 \times 10^{-9} \times n^2$美元，误差=10.91%
  - *n* = the Monthly Active Users (MAUs)
- Facebook与Tencent的年报数据(2003-2020)，适配立方公式
  - 脸书网络的价值（年收入）= $4.20 \times 10^{-18} \times n^3$美元，误差=2.55%；
  - 腾讯网络的价值（年收入）= $4.58 \times 10^{-18} \times n^3$美元，误差=8.48%。

- 实际数据更匹配立方公式
  - 可能暗示着这两个社交网络的价值已经不限于利用节点和边
  - 而是开始利用子集群组，向里德定律过渡

Social networks example
脸书网与腾讯网实例

- Revenue/MAU：每个用户贡献的收入呈现指数增长发展趋势



3. 职业素养：负责任的计算 responsible computing

- Ideas and practices to design and use computing products and services responsibly
  - Cybersecurity issues 安全
  - Privacy awareness 隐私
  - Professional norms 职业规范、职业操守

- Why bother?
- Computing has beneficial and harmful impact to society
  - 计算有正面和负面的社会影响

3.1 Cybersecurity issues 网络空间安全

- The global Internet is under constant attacks
  - Cause harm to society
  - 全球互联网随时被攻击，危害社会

- Example study 危害大，且快速增长
  - McAfee (2020): The Hidden Costs of Cybercrime
    - Cybercrime costed companies worldwide US$1 trillion
    - > 1% of global GDP
    - Was about US$500 billion in 2016

- Compare these to the worldwide computing market
  - The global ICT market: US$3.4 trillion in 2016
  - The global digital economy: US$11.5~24 trillion in 2016
  - The global cybercrime cost: US$0.5 trillion in 2016

Cybersecurity issues

- Cybersecurity problems involve hardware, software and people 网络空间安全涉及硬件、软件、人
  - Not only software such as computer viruses，尽管软件似乎更明显
- Cyber attack types 存在多种攻击，不都是软件侵入
  - Malware: malicious software enabling an attacker to damage or gain unauthorized access to a computer 恶意软件
    - Computer viruses, worms, Trojan horses and spyware 病毒、蠕虫、木马、间谍软件
- 软件故障（bugs）是不是恶意软件（malware）？

Cybersecurity issues

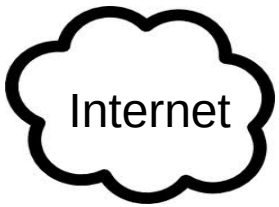
- Cybersecurity problems involve hardware, software and people 网络空间安全涉及硬件、软件、人
  - Not only software such as computer viruses，尽管软件似乎更明显
- Cyber attack types 存在多种攻击，不都是软件侵入
  - Malware: malicious software enabling an attacker to damage or gain unauthorized access to a computer 恶意软件
    - Computer viruses, worms, Trojan horses and spyware 病毒、蠕虫、木马、间谍软件
  - An attack does not have to be in a software form
    - Hardware exploitation 利用硬件的攻击
      - Meltdown: exploiting “out-of-order execution”, a feature of processor hardware 利用硬件的乱序执行
        - Enabling an attacker to read privileged information passwords

Cybersecurity issues

- Cybersecurity problems involve hardware, software and people 网络空间安全涉及硬件、软件、人
  - Not only software such as computer viruses，尽管软件似乎更明显
- Cyber attack types 存在多种攻击，不都是软件侵入
  - Malware: malicious software enabling an attacker to damage or gain unauthorized access to a computer 恶意软件
    - Computer viruses, worms, Trojan horses and spyware 病毒、蠕虫、木马、间谍软件
  - An attack does not have to be in a software form
    - Hardware exploitation 利用硬件的攻击
      - Meltdown: exploiting “out-of-order execution”, a feature of processor hardware 利用硬件的乱序执行
        - Enabling an attacker to read privileged information passwords
  - An attack does not have to install anything on the targeted system
    - Denial-of-service (DoS) attacks, distributed denial-of-service (DDoS) attack 拒绝服务攻击
    - Spams: unwanted emails 垃圾邮件
    - *Phishing*: phishing websites or phishing emails 钓鱼邮件、钓鱼网站

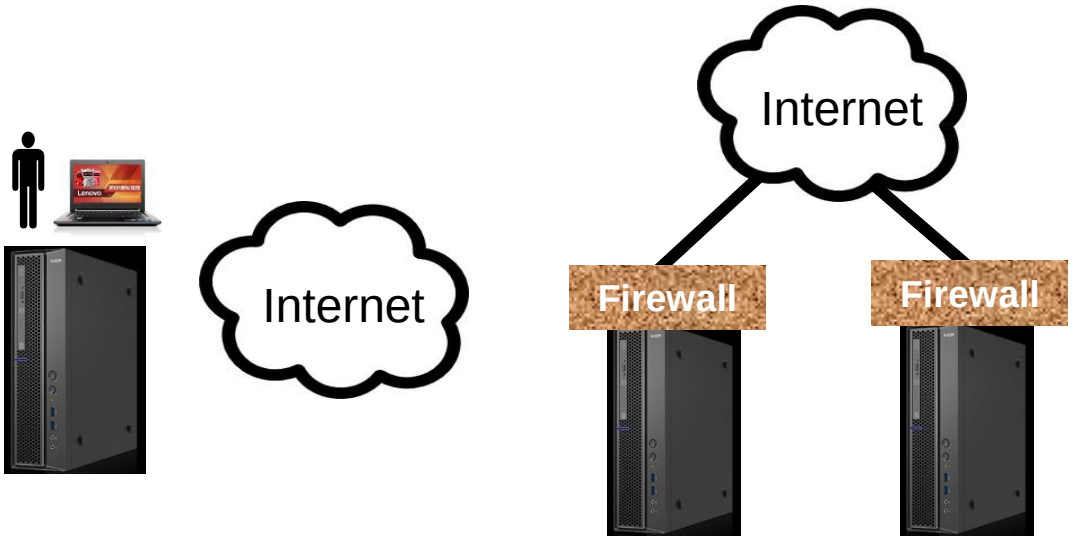
Counter measures 安全措施

- Physical isolation: critical computing systems disconnected from the Internet
- 物理隔离



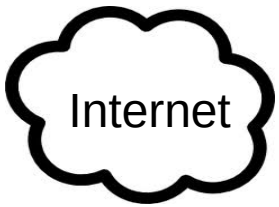
Counter measures

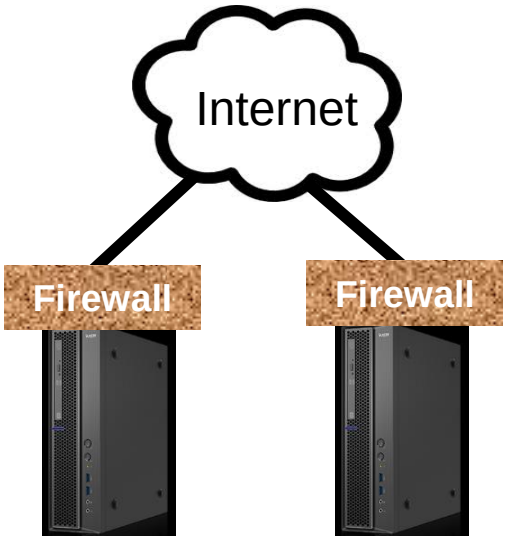
- Physical isolation: core computing systems disconnected from the Internet
- Firewalls: block or filter out undesirable messages
  防火墙

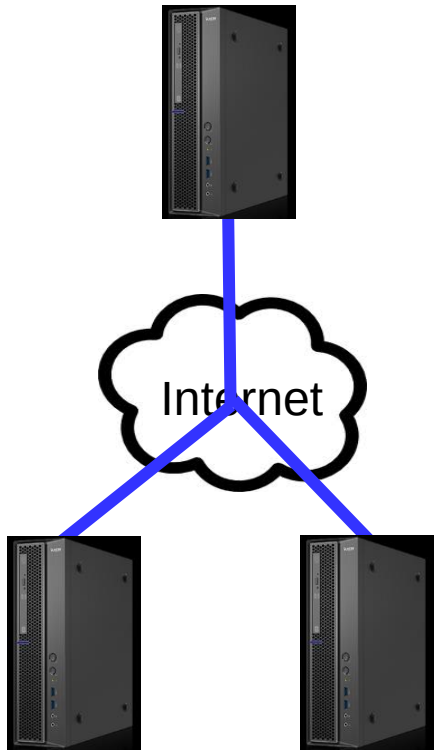


Counter measures

- Physical isolation: core computing systems disconnected from the Internet
- Firewalls: block or filter out undesirable messages
- Virtual private networks (VPNs)
  虚拟私有网







Counter measures

- Physical isolation: core computing systems disconnected from the Internet
- Firewalls: block or filter out undesirable messages
- Virtual private networks (VPNs)

- Antivirus software: detect and kill computer viruses 防病毒软件、杀毒软件

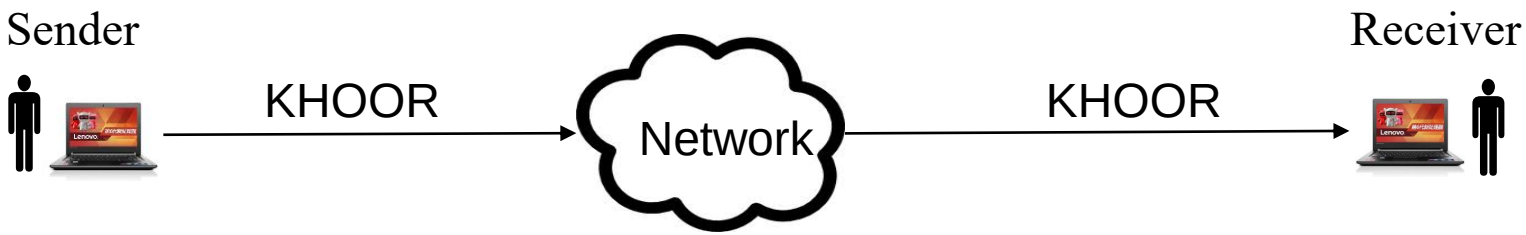
- Cryptography 密码学
  - Secure message communication in the presence of adversaries
  - Encryption: plaintext → ciphertext HELLO → KHOOR 加密：明文→密文
  - Decryption: ciphertext → plaintext KHOOR → HELLO 解密：密文→明文

Symmetric-key encryption: Caesar cipher
对称加密：发送方与接收方共享秘钥；凯撒密码

- Sender and receiver share a key (3 in this example) 此例中秘钥是3
  - **Only a single key is used by both parties, thus symmetric**
- Sender encrypts the plaintext (string of capital letters)
  - By shifting each letter L 3 positions down the alphabet, i.e., ASCII(L)+3
  - E.g., 'H'+3 = 72+3 = 75 = 'K'

  and sends the ciphertext over the network to the receiver
- Receiver decrypts the ciphertext
  - By shifting each letter L up 3 positions, i.e., ASCII(L)-3
  - E.g., 'K'-3 = 75-3 = 72 = 'H'

Alphabet
A 65
B 66
C 67
D 68
E 69
F 70
G 71
H 72
I 73
J 74
K 75
…



HELLO → KHOOR
Shift 3 positions down
Encryption

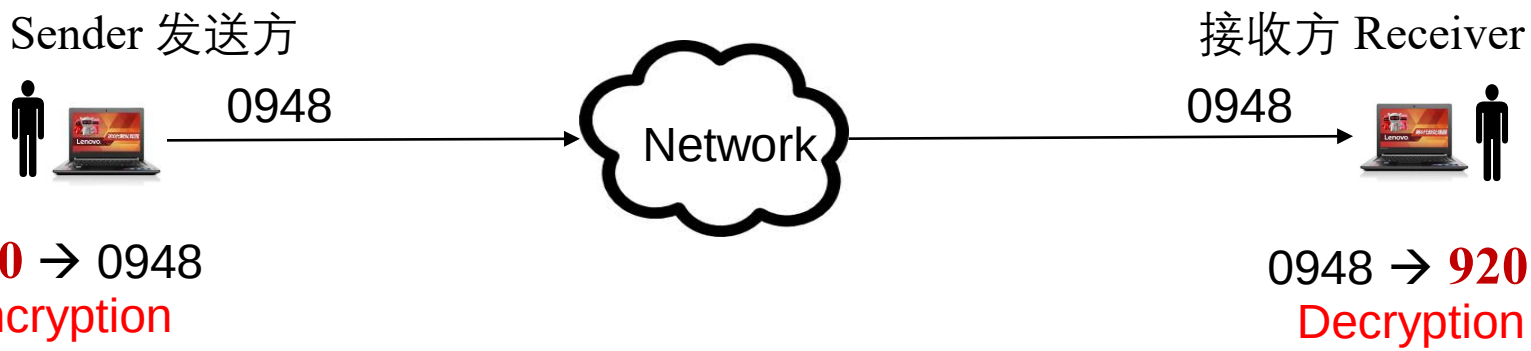
KHOOR → HELLO
Shift 3 positions up
Decryption

***Public-key encryption: the RSA method 公钥加密的RSA方法

- Receiver has two keys
  - 公钥 Public key K_P : known to everybody, including the eavesdropper
    - Used by the sender to encrypt plaintext into ciphertext
  - 私钥（密钥）Private key K_S : known only to receiver; also called **secrete key**
    - Used by the receiver to decrypt ciphertext into plaintext

***Public-key encryption: the RSA method 公钥加密的RSA方法

- Process of securely communicating a plaintext decimal number **920**
  - Receiver makes the magic assumption: *n*=2773,*d*=157,*e*=17
  - Sender
    - Knows the public key $\boldsymbol{K_P} = (e, n) = (17, 2773)$
    - Uses encryption algorithm $C = M^e \bmod n$ to obtain ciphertext *C* from plaintext *M*
      $$C = M^e \bmod n = \mathbf{920}^{17} \bmod 2773 = 948 = 0948$$
    - Sends ciphertext 0948 over the open Internet to receiver
  - Receiver
    - Knows both $\boldsymbol{K_P} = (e, n) = (17, 2773)$ and $\boldsymbol{K_S} = (d, n) = (157, 2773)$
    - Uses decryption algorithm $C = M^e \bmod n$ to obtain plaintext *M* from ciphertext *C*
      $$M = C^d \bmod n = 948^{157} \bmod 2773 = \mathbf{920}$$

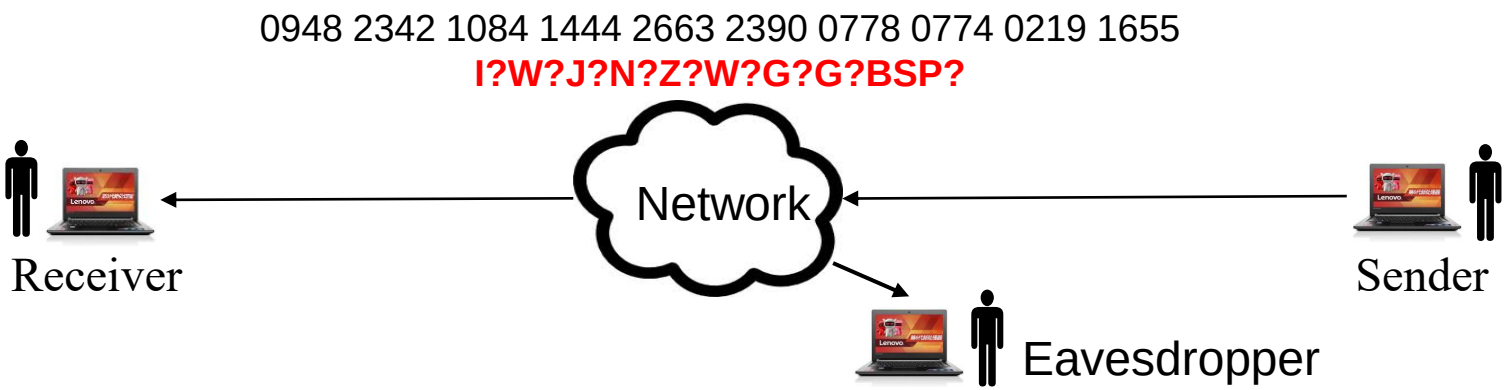


Securely communicating a message

- The plaintext message
  - A 20-character message “ITS ALL GREEK TO ME ”
- Process
  - Sender
    - Encode the text message by: space=00, A=01, B=02, …, Z=26 to obtain a 40-digit number
      - 0920190001121200071805051100201500130500

Securely communicating a message

- The plaintext message: "**ITS ALL GREEK TO ME** "
- Process
  - Sender
    - Encode the text message by: space=00, A=01, B=02, …, Z=26 to obtain a 40-digit number
      - 0920190001121200071805051100201500130500
    - Divide into 4-digit groups
      - 0920 1900 0112 1200 0718 0505 1100 2015 0013 0500
    - Encrypt plaintext number sequence into ciphertext number sequence
      - 0948 2342 1084 1444 2663 2390 0778 0774 0219 1655
    - Sends this ciphertext number sequence to receiver



Securely communicating a message

- The plaintext message
  - A 20-character message “ITS ALL GREEK TO ME ”
- Process
  - Sender
    - Encode the text message by: space=00, A=01, B=02, …, Z=26 to obtain a 40-digit number
      - 09201900011212000718050511002015001305 00
    - Divide into 4-digit groups
      - 0920 1900 0112 1200 0718 0505 1100 2015 0013 0500
    - Encrypt plaintext number sequence into ciphertext number sequence
      - 0948 2342 1084 1444 2663 2390 0778 0774 0219 1655
    - Sends this ciphertext number sequence to receiver
  - Receiver
    - Decrypt ciphertext number sequence into plaintext number sequence
      - 0920 1900 0112 1200 0718 0505 1100 2015 0013 0500
    - Decode number sequence into character string
      - “ITS ALL GREEK TO ME ”

How are the magic numbers determined?

- Magic numbers: n = 2773, d = 157, e = 17

- Process by receiver
  - Randomly choose two large prime numbers p and q, and set $n = p \times q$
    - $p = 47,\ q = 59,\ n = p \times q = 47 \times 59 = 2773$

How are the magic numbers determined?

- Magic numbers: *n* = 2773, *d* = 157, *e* = 17

- Process by receiver
  - Randomly choose two large prime numbers *p* and *q*, and set $n = p \times q$
    - $p = 47,\ q = 59,\ n = p \times q = 47 \times 59 = 2773$
  - Compute the Euler number $(p-1) \times (q-1)$
    - $(p-1) \times (q-1) = 46 \times 58 = 2668$

How are the magic numbers determined?

- Magic numbers: *n* = 2773, *d* = 157, *e* = 17

- Process by receiver
  - Randomly choose two large prime numbers *p* and *q*, and set $n = p \times q$
    - $p = 47,\ q = 59,\ n = p \times q = 47 \times 59 = 2773$
  - Compute the Euler number $(p-1) \times (q-1)$
    - $(p-1) \times (q-1) = 46 \times 58 = 2668$
  - Randomly choose a large integer *d* such that $\mathrm{GCD}(d, 2668) = 1$
    - Set $d = 157$ which satisfies $\mathrm{GCD}(157, 2668) = 1$
    - Complete private key information: $K_S = (d, n) = (157, 2773)$

How are the magic numbers determined?

- Magic numbers: *n* = 2773, *d* = 157, *e* = 17

- Process by receiver
  - Randomly choose two large prime numbers *p* and *q*, and set $n = p \times q$
    - $p = 47,\ q = 59,\ n = p \times q = 47 \times 59 = 2773$
  - Compute the Euler number $(p-1) \times (q-1)$
    - $(p-1) \times (q-1) = 46 \times 58 = 2668$
  - Randomly choose a large integer *d* such that $\mathrm{GCD}(d, 2668) = 1$
    - Set $d = 157$ which satisfies $\mathrm{GCD}(157, 2668) = 1$
    - Complete private key information: $K_S = (d, n) = (157, 2773)$
  - Find value *e* satisfying $(d \times e) \bmod 2668 = 1$
    - $e = 17$ which satisfies $(157 \times 17) \bmod 2668 = 1$
    - Complete public key information: $K_S = (e, n) = (17, 2773)$

RSA allows eavesdropper to know a lot

- A lot of information is open to the world to know
  - The encryption algorithm $C = M^e \bmod n$
  - The decryption algorithm $M = C^d \bmod n$
  - The public key $K_P = (e, n) = (17, 2773)$
  - The character converting scheme: space=00, A=01, B=02, …, Z=26
  - The ciphertext number sequence
    - 0948 2342 1084 1444 2663 2390 0778 0774 0219 1655

RSA allows eavesdropper to know a lot

- A lot of information is open to the world to know
  - The encryption algorithm $C = M^e \bmod n$
  - The decryption algorithm $M = C^d \bmod n$
  - The public key $K_P = (e, n) = (17, 2773)$
  - The character converting scheme: space=00, A=01, B=02, …, Z=26
  - The ciphertext number sequence
    - 0948 2342 1084 1444 2663 2390 0778 0774 0219 1655
- Yet, the eavesdropper cannot decipher the message
  - He lacks the private key $K_P = (d, n) = (157, 2773)$
  - He does not know $d = 157$, which is the solution to $\mathrm{GCD}(d, 2668) = 1$
  - He does not know 2668, which is the Euler number $(p-1) \times (q-1)$
  - He knows $n = p \times q$, but does not know the prime numbers $p,\ q$

RSA allows eavesdropper to know a lot

- A lot of information is open to the world to know
  - The encryption algorithm $C = M^e \bmod n$
  - The decryption algorithm $M = C^d \bmod n$
  - The public key $K_P = (e, n) = (17, 2773)$
  - The character converting scheme: space=00, A=01, B=02, …, Z=26
  - The ciphertext number sequence
    - 0948 2342 1084 1444 2663 2390 0778 0774 0219 1655
- Yet, the eavesdropper cannot decipher the message
  - He lacks the private key $K_P = (d, n) = (157, 2773)$
  - He does not know $d = 157$, which is the solution to $\mathrm{GCD}(d, 2668) = 1$
  - He does not know 2668, which is the Euler number $(p-1) \times (q-1)$
  - He knows $n = p \times q$, but does not know the prime numbers $p,\ q$
- Can the eavesdropper find an efficient algorithm
  - Which recovers prime numbers $p,\ q$?
- Not likely

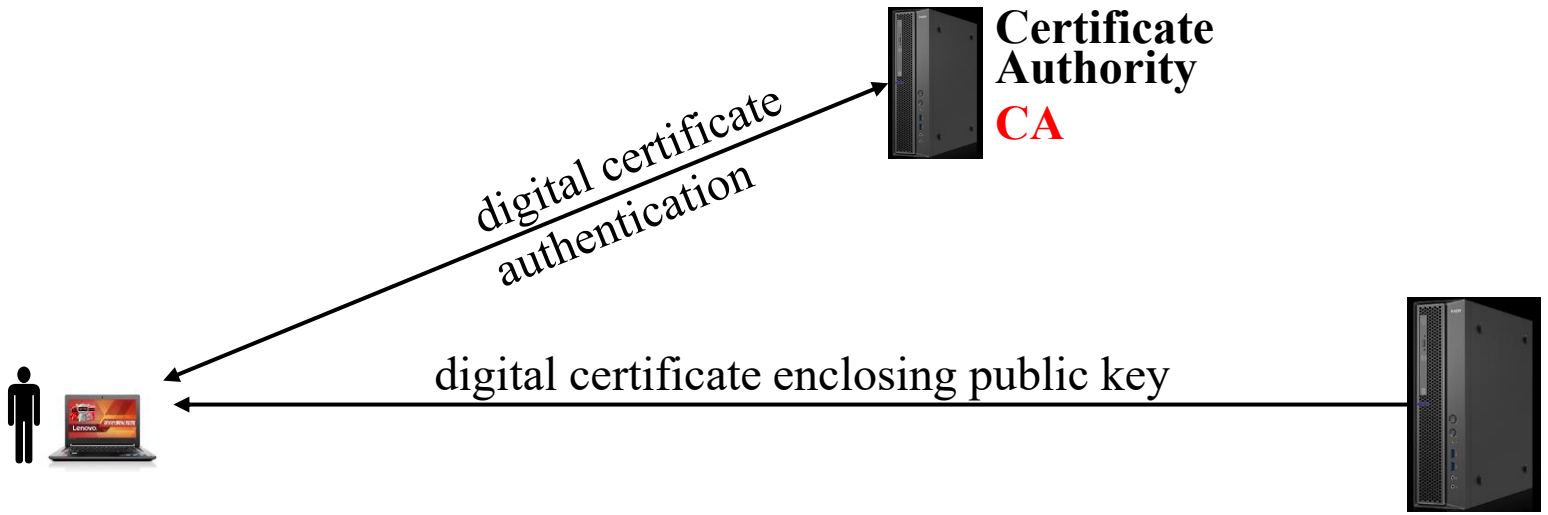
The prime factorization problem

- Given a large natural number n, find the prime numbers $p,\ q$ such that $n = p \times q$
  - Given $n = 2773$, find $p = 47,\ q = 59$, such that $p \times q = 2773$
- This problem has no known efficient algorithm
- RSA relies on this fact

- As of year 2020, the largest RSA integer factored is RSA-250, which has 250 decimal digits
  - A French-US team accomplished the prime factorization task utilizing a network of parallel computers in Europe and the USA
  - The total computing resources used are roughly 2700 core-years
  - At least hundreds of years of computing on a student's laptop

HTTPS: RSA application

- HTTPS = HTTP + Transport Layer Security (TLS)
  - For secure communication between a browser and a website
  - Use symmetric-key and public-key encryption techniques
    - For the long term, use public-key encryption
    - For the short term, use onetime symmetric-key encryption
      - E.g., a HTTP GET session



3.2 Privacy issues 隐私

- Privacy: keeping a user's identity and personally identifiable information (PII) *private*.
- Personal information 个人=自然人
  - Any information relates to a natural person's identity
  - Includes personally identifiable information (PII)
    可区分、追溯到自然人的信息
  - Does not include anonymized personal information
  - 以电子或者其他方式记录的与已识别或者可识别的自然人有关的各种信息，不包括匿名化处理后的信息
- Personal information is broad
  - Such as personal names, ID numbers, personal photos or videos, website clicks records, voice signals, financial records, medical data
- Personal data can be revealed by technology
  - Utilizing metadata, data mining, AI

Sources of further information

- In the computing field
  - *IEEE Security and Privacy* is a professional magazine exploring security and privacy issues
  - Tim Berners-Lee's Solid initiative
- In the legal field
  - GDPR: European Union enacted a law framework, called *General Data Protection Regulation* 通用数据保护条例
    - Went into effect in 2018
  - 中华人民共和国个人信息保护法
    - 2021年11月1日起施行

Basic principles of the laws

- Facilitate protection as well as use of personal information
  - 兼顾个人信息的保护与使用
- A person has basic rights to his/her personal information, such as:
  - Right to permit a third party to collect and use personal data
  - Right to timely rectification of personal data
  - Right to be forgotten
  - Right to port one's personal data to another website
- These rights are protected by law, even when a piece of personal data is not owned by the person
  - A person's cellphone number is protected, even though the number belongs to the telecom company, and the person only "rents" it
- Another person or institution can collect, store, process, and otherwise use a person's data in a legal and fair way
  - PIPA：遵循合法、正当、必要和诚信原则

3.3 Professional norms

- ACM code of conduct: seven principles
  国际计算机协会行为规范的七原则
  - Contribute to society and to human well-being, acknowledging that all people are stakeholders in computing.
    为社会和人类的幸福做出贡献，承认所有人都是计算的利益相关者
  - Avoid harm. 避免伤害
  - Be honest and trustworthy. 诚实可靠
  - Be fair and take action not to discriminate. 做事公平，采取行动无歧视
  - Respect the work required to produce new ideas, inventions, creative works, and computing artifacts.
    尊重他人工作，该工作产生新想法、新发明、创造性作品和计算工件
  - Respect privacy. 尊重隐私
  - Honor confidentiality. 尊重保密协议

https://www.acm.org/code-of-ethics

Form your own thoughtful judgement
了解ACM规范，形成自己的判断

- Understand the ACM code of conduct
  - You don't have to agree to it completely
    - The ACM code itself is evolving
  - But should try to understand what it says
- Apply it to the three examples in textbook, and form your own thoughtful judgement
  应用到教科书三个例子：Examples 67, 68, 69
  - Free flow versus professionally sharing of scientific data
    - 科学数据应该**自由流动**还是**专业性分享**？
      全球共享流感数据倡议组织（GISAID）的规范
  - Full disclosure versus responsible disclosure
    - 组织内部出现可能有害社会的漏洞，应该向社会完全曝光还是负责任地通报
  - The case of the Morris worm
    - 善意的（无恶意的）科学研究工作是否可以越界？