

Compsci 82, Fall 2009, Takehome Final

You must complete this via blackboard. These questions are shown here for you to study before taking the online, takehome final. You may use books, notes, and the Internet. You **may not talk to any person about these questions except for Prof. Astrachan**

Complete the questions on Blackboard, the questions are shown here only for studying.

Question 1

1 points

In Internet measurement tests the average size of an IP packet is fairly close to 512 bytes. That's 2^9 bytes, or 0.5 Kb (half a kilobyte where a Kb is 2^{10} bytes, about a thousand bytes). The average size of a Netflix or iTunes streamed HD movie is about 4 Gb --- that's 4 Gigabytes where a Gigabyte is 2^{30} bytes or about a billion bytes.

If all packets are average, about how many packets are needed to stream an average HD movie? Note that a billion is a thousand million, and that 2^{20} is about a million. Remember that $2^{13} = 2^3 \times 2^{10}$ so 2^{13} is about 8 thousand --- since $2^3 = 8$ and 2^{10} is about a thousand ($2^{13} = \text{exactly } 8,192$).

- ☐ About two hundred thousand packets
- ☐ About eight hundred thousand packets
- ☐ About two million packets
- ☐ About eight million packets

Question 2

1 points

Most estimates indicate that the number of stars in the universe (not just our galaxy, but the universe) is about 10^{22} though this isn't exact. Using this estimate, about how many IPv6 addresses could be allocated to each star in the universe?

- ☐ About 10^8
- ☐ About 10^{16}
- ☐ About 10^{24}
- ☐ About 10^{32}

Question 3

1 points

More than 35 million iPhone and iPod touch devices have been sold. There are more than 350 million Facebook users and 50% of them login on a given day. Using these estimates which one of the following is **LEAST LIKELY** to be true?

- ☐ The total number of apps used by all iPhone/touch users in a single day is more than the number of IP packets sent or received by Facebook users

on the same day.

- ☐ The total amount of money spent by all iPhone users as part of their cell-phone service on a given day is more than the amount of money spent by all Facebook users to use Facebook on the same day.
- ☐ More people use Facebook from a laptop or desktop machine than there are people accessing Facebook from their iPhone on a given day.
- ☐ If every time someone downloads an app to an iPhone (even free apps) Apple makes \$0.01 in revenue, and each time someone logs into Facebook generates \$0.01 in revenue for Facebook, then Facebook has earned more money through this process in the last three years than Apple has earned from this download fee.

Question 4

1 points

A 10 megapixel camera stores an image in raw format so that the image takes roughly 10 MBytes to store (that's 10 million bytes, a byte is 8 bits). Mobile-phone WiMax networks support upload speeds of 2Mbps, that's 2 Million bits/second. How long will it take to upload one raw image from a 10 megapixel camera over a WiMax network?

- ☐ About 40 seconds
- ☐ About 2 minutes
- ☐ About 20 minutes
- ☐ About 4 hours

Question 5

1 points

When Christopher Poole, aka moot, visited our class, which one of the following was **NOT** something he told us about 4chan users?

- ☐ The 4chan site has contributed several memes to Internet culture including rickrolling and lolcats.
- ☐ There are more than six million visitors to the 4chan site each month and more than 400 million page views per month.
- ☐ The archiving costs to store all the images uploaded to 4chan over time constitute most of the money spent on running 4chan.
- ☐ The icanhascheezburger.com site was bought for roughly two million dollars, but moot doesn't see any of that money since it's not his site.

Question 6

1 points

When Duke alum Luis von Ahn visited our class, which one of the following was **NOT** part of the information/message he conveyed to the class?

- ☐ Luis has been offered money for the software and projects he has built, including reCaptcha and the image labeler, but his respect for academic freedom and his lifelong desire to be a professor has kept him from selling any of his projects or intellectual endeavors.

- ☐ His reCaptcha project was designed to leverage potentially non-productive time, e.g., time spent verifying you're human and not a bot, to do useful and worthwhile things.
- ☐ People play the games he has developed, so-called *Games With a Purpose*, mostly because they're fun to play rather than because they're doing something worthwhile.
- ☐ Luis' next project involves teaching people a language they don't know by having them translate documents written in that language into a language they do know.

Question 7**1 points**

In class an excerpt from a movie was shown to buttress the idea of freedom associated with the free and open source software movements, especially the idea that freedom is very, very important. Which movie and actor were the source of the 'freedom' excerpt?

- ☐ Bruce Willis in *Live Free or Die Hard*
- ☐ Mel Gibson in *Brave Heart*
- ☐ Angelina Jolie in *Hackers*
- ☐ Ben Stein in *Ferris Bueller's Day Off*

Question 8**1 points**

The number of cookies served by the website TMZ.com is best characterized by which of the following?

- ☐ Fewer than 20 cookies.
- ☐ More than 20, but fewer than 30 cookies
- ☐ More than 30 cookies, but fewer than 100 cookies.
- ☐ More than 100 cookies.

Question 9**1 points**

The website bit.ly is used by twitter and other sites to shorten lengthy URLs. For example, the URL

<http://www.cs.duke.edu/courses/cps082/fall09/syllabus.php>

is shortened by bit.ly to the URL below

<http://bit.ly/8Yt2Su>

The six-character URL *8Yt2Su* is composed of uppercase and lowercase letters and the digits 0-9. About how many different six-character URLs are there composed of uppercase letters, lowercase letters, and the digits 0-9?

- ☐ About 14 times as many as there are IPv4 addresses.
- ☐ About as many as there are atoms in the universe.
- ☐ About half as many as there are IPv6 addresses

- ☐ About one million.

Question 10**1 points**

Which one of the following is **NOT** distributed under a free or open source software license?

- ☐ Adobe's reader software for viewing PDF documents.
- ☐ The Apache webserver software.
- ☐ The MySQL database software.
- ☐ The Linux operating system software.

Question 11**1 points**

In an article about Oracle trying to buy MySQL from here

http://news.cnet.com/8301-13505_3-10379280-16.html

We find the following:

The GPL, which is supposed to be the ultimate guarantor of software freedom, may deliver the opposite. Because of its control-freak urges, it can stymie competition, which is presumably why Stallman is now calling on the European Commission to grant what his license couldn't: freedom.

...

Apache leaves everyone--developers, users, vendors, etc.--on equal footing. The GPL does not. With the GPL, the copyright holder retains effective control. That's one reason it [the Apache License] has been so popular with commercial open-source companies, but the Oracle/MySQL situation may prompt more companies to consider using an Apache license so as to preserve maximum freedom in case of takeover, hostile or otherwise.

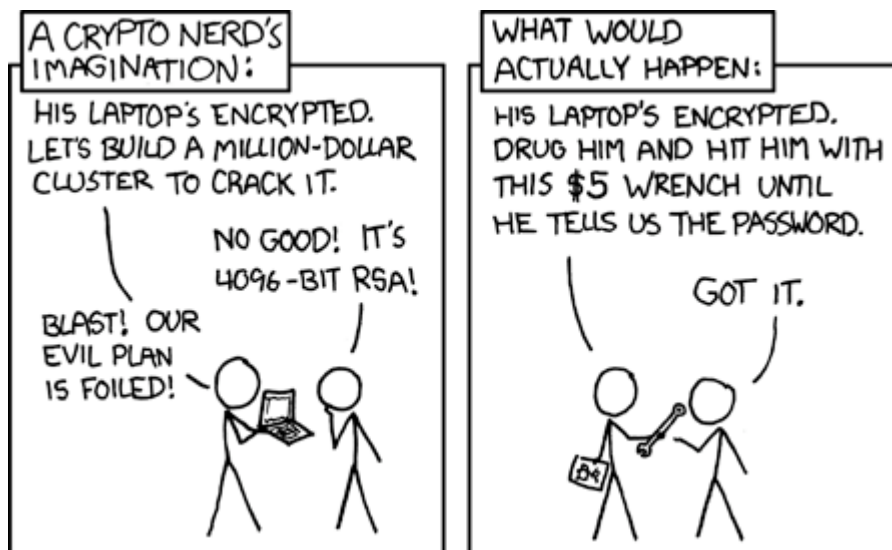
Which of the following is LEAST likely to be a reason to prefer the Apache license to the GPL license?

- ☐ Larry Ellison, the CEO of Oracle, is a billionaire, whereas Richard Stallman, the developer of the GPL, is not. The Apache license helps bridge the gap between rich and poor.
- ☐ The GPL is a copyleft or viral license -- it essentially forces all derived software to be distributed using the same license, making it possibly difficult to commercialize.
- ☐ The Apache license allows derived software to be distributed under different licenses as long as the new software acknowledges the Apache licensed code -- thus new software could even be closed source with an acknowledgment, making it perhaps easier to commercialize.

- There are often subtle and nuanced differences between licenses, open or closed source. The differences between the Apache and GPL licenses may be straightforward to a non-legally oriented view, but companies must be sure that their software is licensed appropriately.

Question 12**1 points**

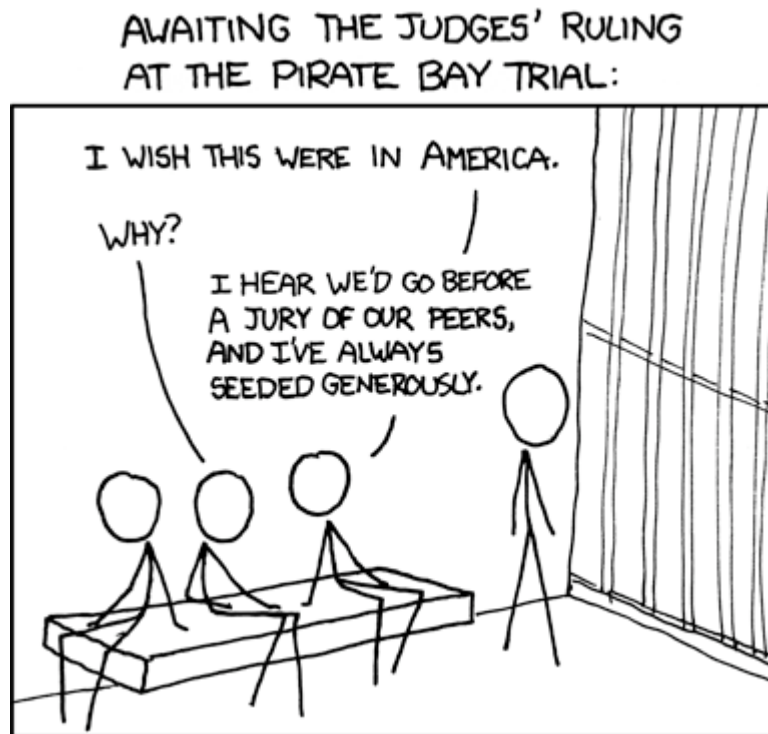
Which of the following best explains the 'No Good! ...' response in the left panel of the XKCD comic below.



- RSA stands for 'Really Strong Algorithm' ensuring the encryption can't be broken.
- A 4,096 bit encryption scheme is strong because there are too many combinations represented by 4096 bits making it impossible to break such encryption using brute force, computational techniques.
- 4096 is 2^{12} and cryptographers and nerds both embrace the idea of 'a baker's dozen' different kinds of approaches to cryptography -- since 12 is a dozen, and a baker would have 'bread' or money, this is a setup for the second panel of the comic.
- The phrase '4096-bit RSA' is leet-speak for using a shield to protect against gamma rays that are emitted from most laptops --- this makes the next panel very funny because of the phrase 'got it' which is leet-speak for moron.

Question 13**1 points**

Which of the following best explains the humor in the comic below.



- ☐ Pirate Bay is a website that was used to store Bittorrent information; Bittorrent is a peer-to-peer network; and bittorrent clients had to help upload as well as download, a process known as 'seeding'.
- ☐ Pirate Bay was a website in Denmark used for peer-to-peer network information and the Danish system of trial-by-jury uses a philosophy known as 'seeding to ensure growth of information'.
- ☐ The Pirate Bay trial refers to a trial of some of the prisoners kept in Guantanamo Bay, which is part of Cuba and in an area that was frequented in the past by pirates. The prisoners wish they had a trial based on a jury of peers rather than a military trial which would be likely to result in a harsher more severe punishment.
- ☐ The seated stick figures represent the three fates from Greek mythology that spun the thread of life and who could 'peer' or 'see' into the future of a person's life. The Pirate Bay Trial is a subtle allusion exploiting the similarities in the word 'pirate' and 'private' which differ only by the letter 'V' for victory.

Question 14**1 points**

Which of the following is/are true about Bittorrent and its use?

- I. A file downloaded via a Bittorrent client will download more quickly if the person downloading also shares part of the file for uploading by others.
- II. The bittorrent protocol is not protected by patent or by other intellectual property constraints --- anyone can write a bittorrent client if they have the ability and the desire without fearing legal sanctions from Bram Cohen, the developer of the protocol.

III. Bittorrent files can download more quickly than other p2p networks because the downloading is spread cross multiple bittorrent users, not from a single user as is the case with Kazaa, Limewire and most p2p clients.

- ☐ I only
- ☐ I and II only
- ☐ II and III only
- ☐ I, II, and III

Question 15**1 points**

Which of the following is/are true about SPAM (unsolicited email)

I. SPAM is often delivered by botnets: thousands to millions of machines that have been corrupted by viruses/malware etc., so that they can help deliver email/SPAM when given commands to do so by command/control software run by groups that rent their services to spammers.

II. Sending SPAM that appears to come from a known friend/user is possible because many parts of an email message header can be faked or forged.

III. SPAM is not rewarding in any economic sense. SPAM is only sent to annoy users and to show it's possible by hackers and groups eager to show that they can be annoying to normal/regular users.

- ☐ I only
- ☐ II only
- ☐ I and II only
- ☐ II and III only
- ☐ I, II, and III

Question 16**1 points**

Which of the following is/are true regarding Internet censorship?

I. Internet users can use TOR, a project initially sponsored by the U.S. Navy and government, to bypass restrictions imposed by country-wide firewalls or censors.

II. China succeeds in censoring the Internet in large part because although the Chinese population is more than a billion, virtually no one in China is online, so filtering and censorship is not logistically difficult.

III. Software created by a Chinese group, Falun Gong, is used to send links via email that permit users to bypass filters based on IP addresses, because the links rediret the user to a filtered IP address through another IP address.

- ☐ I only
- ☐ I and II only

- ☐ II and II only
- ☐ I and III only
- ☐ I, II, and III

Question 17**1 points**

When Duke alum Ge Wang Skyped into class to discuss his work, which of the following took place?

I. He indicated he had never used Skype before and "Skype rocks dude!!"

II. He used the phrase "it depends" in his powerpoint slides as part of his presentation.

III. He sang to the class in demonstrating the "I Am T-Pain" iPhone app his company developed.

- ☐ I only
- ☐ II only
- ☐ I and II only
- ☐ II and III only
- ☐ I, II, and III

Question 18**1 points**

Thomas Narten spoke in class bout issues with IPv4 and why we should switch to IPv6. Which of the following is/are reasons he mentioned as a reason to switch?

I. Many people use NAT (Network Address Translation) to cope with the need for multiple devices to share on actual IP address. But NAT has problems, often it's at fault when the Internet doesn't appear to work from a consumer perspective.

II. A market for blocks of IP addresses might develop that would/could cause problems with both engineering and social aspects of the Internet. Its possible such a marketplace could be illegal, but it would still happen if some users/groups were advantaged by it even if it meant trouble from an engineering perspective to swap blocks of addresses.

III. It's possible to make a switch to IPv6 without doing away with IPv4 --- the two protocols can co-exist making a transition more possible than it might first appear to be.

- ☐ I only
- ☐ III only
- ☐ II and III only
- ☐ I and II only

- ☐ I, II, and III

Question 19**1 points**

In class we saw a comparison of the sites and images returned by google.cn (The Chinese Google site) and google.com (the U.S. Google site) using this website: <http://www.computerbytesman.com/google/imagesearch.htm> for the search queries *tianaman* and *tiananmen*. The results returned were very different for the second of these queries, and substantially similar for the first query. Which of the following best explains this observed phenomenon.

- ☐ The Google in China site does not store photos or results of anything that the Chinese Government would view as presenting China in a bad light. The query *tianaman* is misspelled, so the Chinese site shows images from the U.S. site since the Chinese site cannot handle misspelled queries in English.
- ☐ The Google in China site is looking for specific queries and treating them differently, essentially censoring the returned results when a search query matches a list of those for which different results should be returned, like *tiananmen*. Since the query *tianaman* is misspelled, it doesn't get caught by the literal-matching filter that looks for specific key words, thus unfiltered/uncensored results are returned for the misspelling.
- ☐ The English words and spelling of *tiananmen* and *tianaman* are actually translated into completely different words in Chinese. The first translates into 'Gate of Heavenly Peace' which is why the Chinese site shows images of beauty and color whereas the U.S. site doesn't because the English word just matches the square that is the site of an event in 1989 in which hundreds were killed.
- ☐ The Google did-you-mean query fixing mechanism doesn't work in the Chinese Google site, so the query isn't corrected on that site as it is on the U.S. Google site.

Question 20**1 points**

Section 512 of the DMCA and section 230 of the CDA each protects ISPs and other content providers --- essentially providing them with immunity from liability. Which one of the following is/are true related to these laws/statutes?

I. Both sections/laws protect the ISP/content provider in the case when a third party (e.g., not the ISP) posts infringing or defamatory material -- the commonality is the third party, the material may be either infringing (e.g., the DMCA) or defamatory (e.g., the CDA).

II. If John Smith writes original content in a Wikipedia entry that he makes up regarding the sanity of Isabel Jones, in which Mr. Smith claims Ms. Jones is insane for not going out with him, Wikipedia must remove the material at the request of Ms. Jones because of Section 512 of the DMCA, regardless of

whether Ms. Jones is actually insane.

III. The DMCA section and laws are about copyright, whereas the CDA (Communications Decency Act) is about defamation --- so while they both provide protection and immunity, this immunity is in relation to different areas of the law.

- ☐ I only
- ☐ II only
- ☐ III only
- ☐ I and III only
- ☐ I, II, and III

Question 21

1 points

It takes several seconds to load the web page from Tsinghua University in China from www.tsinghua.edu.cn/eng in my browser at Duke. Which of the following is **LEAST** likely to be true when someone down the hall from me loads the same page a minute later.

- ☐ The person down the hall might experience a faster load time because a local DNS server has cached the IP address corresponding to www.tsinghua.edu.cn so that less time is taken in finding the IP address for there browsing session.
- ☐ The person down the hall might experience a slower load time because packets on the Internet are routed differently depending on local traffic, so packets to their machine could take longer to arrive just because of Internet traffic in Los Angeles that might slow things down.
- ☐ The person down the hall might experience a faster load time because all web pages from China are stored in a central server at Duke once they've been retrieved the first time so that they can be scanned for malware -- thus their load is faster because they load from the local, Duke copy of the webpage.
- ☐ It's possible that the person down the hall will experience a slower load time because they are using a wireless network on their laptop while I'm using a wired connection from my desktop machine.

Question 22

1 points

The number of bits in an IPv4 (that's a 4, not a 6) address is

- ☐ 32 bits
- ☐ 64 bits
- ☐ 128 bits
- ☐ 256 bits

Question 23

1 points

Which of the following is the correct chronological ordering of events related to the Internet (arranged from earliest to latest/most recent in terms of the time at which the event happened).

- ☐ Web, Internet, p2p networks, IPv6 deployment
- ☐ Internet, Web, p2p networks, IPv6 deployment
- ☐ IPv6 deployment, p2p networks, Web, Internet
- ☐ Internet, Web, IPv6 deployment, p2p networks

Question 24**1 points**

Which of the following is/are quotes from Richard Stallman. You can reason about this by understanding Stallman's basic philosophy with respect to free software.

I. For personal reasons, I do not browse the web from my computer. (I also have no net connection much of the time.) To look at page I send mail to a daemon which runs wget and mails the page back to me. It is very efficient use of my time, but it is slow in real time.

II. The War on Drugs has continued for some 20 years, and we see little prospect of peace, despite the fact that it has totally failed and given the US an imprisonment rate almost equal to Russia. I fear that the War on Copying could go on for decades as well. To end it, we will need to rethink the copyright system, based on the Constitution's view that it is meant to benefit the public, not the copyright owners. Today, one of the benefits the public wants is the use of computers to share copies.

III. It's clear that other problems such as religious fundamentalism, overpopulation, damage to the environment, and the domination of business over government, science, thought, and society, are much bigger than non-free software. But many other people are already working on them, and I don't have any great aptitude or ideas for how to address them. So it seems best for me to keep working on the issue of free software. Besides, free software does counter one aspect of business domination of society.

- ☐ I only
- ☐ II only
- ☐ III only
- ☐ II and III only
- ☐ I, II, and III

Question 25**1 points**

The IETF protocols SMTP (for mail), HTTP (for web traffic), and BGP (for communication between Internet Autonomous Systems) help different parts of the Internet work. Which of the following is/are true regarding the IETF and the functioning of the Internet using these and other, related protocols designed by the IETF.

I. The last P in HTTP, SMTP, and BGP as well as the P in IP represents the word 'Protocol'.

II. Every protocol designed by the IETF ends in a P to indicate it is a protocol. This is specified by the ISP: the IETF Standard Protocol.

III. IETF meetings are open to the public, but attendance at a meeting is not required to participate in extensive email and list discussions regarding IETF proposals and protocols.

- ☐ I only
- ☐ III only
- ☐ I and II only
- ☐ I and III only
- ☐ I, II, and III