

Answer **ALL** questions. Write your answers **on this exam paper**.

Write your name on this exam paper in the box below, and your initials at the top right corner of all other pages of this exam paper.

Name:

Marks for each part of each question are indicated in square brackets

Calculators are permitted

Part ONE

1. Multidimensional parity checking

As described in K&R, one way of detecting errors is to transmit data as a block of m rows of n bits per row ($d_{i,j}$), adding parity bits to (*i.e.* modulo-2 sums of) each row (p_i) and column (q_j). The lower-right corner bit r is the modulo-2 sum of parity bits p_1, \dots, p_m .

$$\begin{array}{cccc|c} d_{1,1} & d_{1,2} & \cdots & d_{1,n} & p_1 \\ d_{2,1} & d_{2,2} & \cdots & d_{2,n} & p_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ d_{m,1} & d_{m,2} & \cdots & d_{m,n} & p_m \\ \hline q_1 & q_2 & \cdots & q_n & r \end{array}$$

- a. Give an example of a four-bit error that would not be detected by the two-dimensional parity scheme above.

[3 marks]

- b. What is the general set of circumstances under which four-bit errors will be undetected?

[3 marks]

- c. Consider the case where $m = n = 2$. Assuming that exactly four bits are errored in a packet, what is the probability that those bit errors go undetected? [3 marks]

[Total for Question 1: 9 marks]

2. UniLeaks

Inspired by the recent success (or notoriety!) of *WikiLeaks* (an international organization that publishes submissions of otherwise unavailable documents from anonymous sources and leaks), you secretly hand-pick a small group of GZ01 students to join you in a similar effort at a University-wide scale, exposing corruption, malfeasance, and wrongdoing among your lecturers.

Given the high impact of your work and your paranoid bent, you decide you can't trust the computer software you use to be free of wiretapping backdoors, so using skills learned in 3005 and GZ03, and a few late, caffeine-fuelled nights, you write your own operating system and encryption routines. Now all that remains is to write software to compress the files sent over the network. You decide to use Lempel-Ziv-Welch (LZW) compression, recalling the pseudocode for LZW presented in lecture:

	<i>Table:</i>
LZW-ENCODE(<i>message</i>)	{
1 <i>curseq</i> ← ""	'a' : 0,
2 for <i>jj</i> ∈ 1.. <i>length(message)</i>	'b' : 1,
3 do if <i>curseq</i> + <i>message</i> [<i>jj</i>] ∈ <i>Table</i>	'c' : 2,
4 then <i>curseq</i> ← <i>curseq</i> + <i>message</i> [<i>jj</i>]	'd' : 3,
5 else Output <i>Table</i> [<i>curseq</i>]	'e' : 4,
▷ Assign the new symbol the	'f' : 5,
▷ next consecutive number:	'g' : 6,
6 <i>Table</i> [<i>curseq</i> + <i>message</i> [<i>jj</i>]] ← size(<i>Table</i>)	'h' : 7,
7 <i>curseq</i> ← <i>message</i> [<i>jj</i>]	'i' : 8,
8 Output <i>Table</i> [<i>curseq</i>]	'j' : 9,
	'k' : 10,
	'l' : 11,
	'm' : 12,
	'n' : 13,
	'o' : 14,
	'p' : 15,
	'q' : 16,
	'r' : 17,
	's' : 18,
	't' : 19,
	'u' : 20,
	'v' : 21,
	'w' : 22,
	'x' : 23,
	'y' : 24,
	'z' : 25,
	}

Here we represent symbols as integers. Assume that LZW-ENCODE uses *Table*, whose initial contents are shown above to the right.

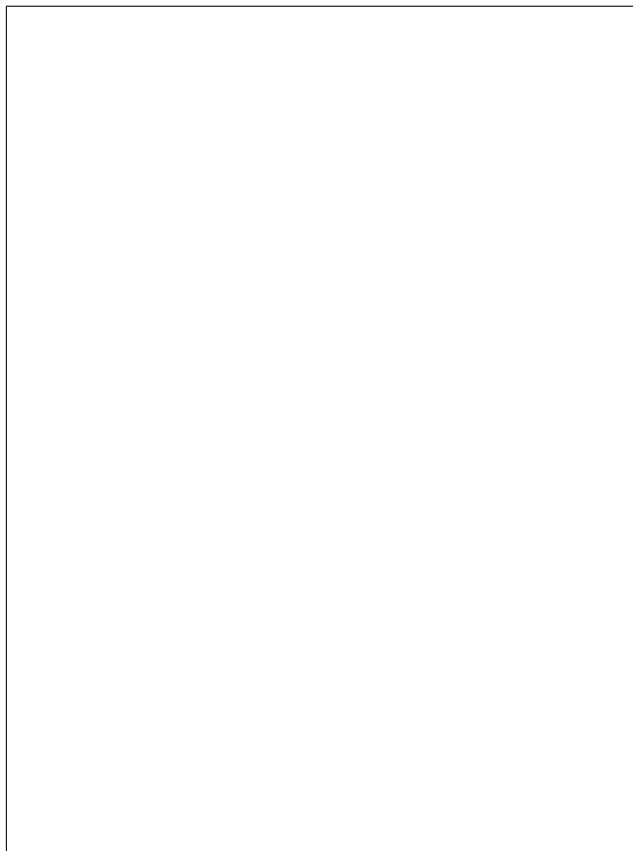
- b. Your classmate Julian notices that because new symbols are added to the sender's table immediately after they are transmitted, the decoder can derive `Table` as it processes the coded message symbols. He claims that it is thus possible to decode the message without the sender transmitting `Table` to the receiver.

Write **pseudocode** for procedure `LZW-DECODE(symbols)`, which takes an encoded message *symbols* as an ordered array of integers, and decodes *message* **without** the encoding table *Table* constructed at the sender.

In this part, you may take *symbols* to be an array of integers, and you may assume the existence of a `+` operator that concatenates strings together and a Python-like substring operator `[]` that extracts parts of strings (*i.e.*, if `a = 'GZ01'` then `a[0] = 'G'`). You may also (as in `LZW-ENCODE`) assume the existence of a `size` operator that returns the number of entries in *InverseTable*.

InverseTable:

`LZW-DECODE(symbols)`



```
{ 0: 'a',
  1: 'b',
  2: 'c',
  3: 'd',
  4: 'e',
  5: 'f',
  6: 'g',
  7: 'h',
  8: 'i',
  9: 'j',
 10: 'k',
 11: 'l',
 12: 'm',
 13: 'n',
 14: 'o',
 15: 'p',
 16: 'q',
 17: 'r',
 18: 's',
 19: 't',
 20: 'u',
 21: 'v',
 22: 'w',
 23: 'x',
 24: 'y',
 25: 'z' }
```

[12 marks]

- c. Recalling the definition for the entropy (information content) H of a discrete source given in lecture:

$$H = \sum_{i=1}^L P(A_i) \log_2 \left(\frac{1}{P(A_i)} \right)$$

and assuming that the transmitted message is an infinitely-repeating sequence $\text{abbabbabb} \dots$, what is the approximate entropy (in bits/letter) of this message? [3 marks]

- d. Assuming that each LZW symbol is transmitted using a different five-bit representation (e.g. $\text{b} = 00001$), how many bits/letter does the LZW encoder transmit for the message in Part (a)? [3 marks]

- e. For the message above in Part (a), which would be a better choice—Huffman coding or LZW—and why? In your answer, be sure to specify the tree the Huffman code would construct, and the bits Huffman coding would send over the wire. [5 marks]

[Total for Question 2: 33 marks]

3. Abusing TCP

Recall that the Transmission Control Protocol (TCP) header includes a flag bit named RST, for “reset.” A TCP segment whose TCP header has the RST bit set is called “an RST” for short. The TCP protocol specification states that when a host receives an RST on a connection for which it holds state, it immediately aborts that connection *if certain conditions are met*. Those conditions are as follows:

- the source IP address, source port, destination IP address, and destination port in the RST match those of a connection known to the receiver of the RST
- the sequence number S on the RST falls within the current receive window for that connection known to the receiver of the RST (*i.e.*, $L \leq S < H$, where L is the next contiguous sequence number expected from the sender, and $H = L + R$, where R is the receiver’s advertised window size)

If these conditions are not met, then the receiver of the RST ignores the RST.

- a. Your Internet Service Provider (ISP), acting at the request of major movie studios, decides to block BitTorrent traffic to and from its customers. (BitTorrent sends data over TCP.) Suppose your ISP monitors all TCP traffic on your DSL line, and searches through the TCP payload for the BitTorrent protocol.

Describe how your ISP can construct an RST that, when sent to your home PC, will immediately kill any BitTorrent TCP connection your home PC has established. In your answer, be sure to describe exactly what values any relevant fields in the RST’s TCP header must have, and **how the ISP determines these values**.

[5 marks]

- b. GCHQ (the UK's communications intelligence agency) learns that a foreign spy in the UK has obtained access to secret nuclear warhead designs, and intends to send these designs (using TCP over the Internet, of course) from a server with IP address S in the UK to a server with IP address D in his home country.

GCHQ needs to act quickly to prevent this file transfer, and does *not* have time to physically install any monitoring or other networking equipment on the physical path between S and D .

An officer at GCHQ who took 3035/GZ01 years ago hopes to prevent this file transfer by sending carefully constructed RSTs to D . The officer has a third Internet-connected host, M , that is *not* on the physical path between S and D .

Assume that S uses a fixed TCP source port known to the GCHQ officer for its connection, and that D listens for TCP connections on a well-known TCP destination port (known to the spy and the GCHQ officer).

Assume further that the default receive window size used by modern TCP receivers is 64 Kilobytes, and that the officer is aware of this.

- i. Which TCP header values necessary for the correct functioning of the RST (*i.e.*, to kill the flow from S to D) does the GCHQ officer know and which does he not know?

[3 marks]

- ii. How many unique RSTs (where "unique" means different in at least one bit in the TCP header) must the officer send from M to D to be *certain* that he has caused the TCP connection from S to D to abort? (Hint: think about the *minimum* number of unique RSTs; it may not be necessary to send every possible unique RST.)

[6 marks]

- iii. Suppose that the Internet path from the GCHQ officer's machine M to D can carry traffic at 10 Mbit/s and the path from S to D can carry traffic at 10 Mbit/s concurrently.

How long will it take the officer to send the necessary number of RSTs from M to D to be certain of killing the connection from S to D ? [3 marks]

- c. An attacker controls an Internet-connected host M . He wishes to cause trouble for the operator of another Internet-connected host P (the *patsy*, *i.e.*, person unjustly blamed for another person's wrong-doing) by causing the appearance that P is flooding a third Internet-connected host A , the *accuser*, with traffic.

Moreover, the attacker wishes to be stealthy and minimize any chance that he might ever be associated with this attack. So he does not want to send *any* traffic from M with an IP destination address of A .

How can M leverage IP source address spoofing and the TCP handshake to send traffic stealthily so that A will believe that P is flooding A with traffic, yet M sends *no* traffic with an IP destination of A ?

[8 marks]

[Total for Question 3: 25 marks]

Part TWO

Multiple Choice

The questions in this section are multiple choice. Zero, one, or more than one choice may be correct, and the number of correct choices varies from question to question. Circle the letters below corresponding to ALL correct choices.

Do not guess. You will be awarded one mark for each choice correctly circled, but you will lose one mark for each choice incorrectly circled. Final results in this section will be normalized across the class.

1. ARP

Which of the following are true statements about the Address Resolution Protocol (ARP)?

- A. ARP resolves IP addresses to MAC addresses for hosts on different subnets.
- B. Most data link layer hardware understands IP addresses.
- C. IP routers must forward ARP messages to ensure reachability between two IP hosts.
- D. The Spanning Tree Protocol must forward ARP queries across blocked ports for complete connectivity in a bridged Ethernet.
- E. The link layer protocol header allows the demultiplexing of ARP packets to the network layer.

2. CIDR

Which of the following are true statements about Classless Inter-Domain Routing (CIDR)?

- A. A network on the Internet with subnet mask 255.255.240.0 can handle a maximum of 2047 hosts.
- B. Assuming they all share the same outgoing link, an IP router can aggregate the following four IP addresses to one /20 address: 57.6.96.0/21, 57.6.104.0/21, 57.6.112.0/21, and 57.6.120.0/21.
- C. 135.46.0.0/16 is an example of a Class B network.
- D. At a CIDR-aware IP router, a packet with destination IP address 200.23.18.1 would match routing table prefix 200.23.18.0/23 before 200.23.16.0/20.
- E. At a CIDR-aware IP router, a packet with destination IP address 200.23.18.1 would match routing table prefix 200.23.18.128/24 before 200.23.18.0/20.

3. DNS

- A. For the purpose of registering its domain name `cs.ucl.ac.uk`, the Computer Science department had to obtain permission only from the `ucl.ac.uk` domain registrar.
- B. Root name servers receive and respond to most recursive queries.
- C. Most DNS queries and responses use UDP.
- D. Kaminsky-style DNS cache poisoning requires the attacker to issue DNS queries for valid names in the domain she intends to compromise.
- E. A DNS name is a type of path name.

4. Ethernet

Which of the following are true statements about the Ethernet as discussed in lecture? Assume that τ denotes one end-to-end propagation delay of a signal on a single-segment, shared Ethernet.

- A. The beginning of stations' backoff windows may differ by at most 2τ .
- B. Given the same number of transmitters waiting, and discounting wireless loss, slotted ALOHA and Ethernet share approximately the same acquisition probability.
- C. Given the same number of transmitters waiting, and discounting wireless loss, unslotted ALOHA and Ethernet share approximately the same acquisition probability.
- D. Ethernet repeaters are Layer-1 devices.
- E. An Ethernet receiver can decode part of a Manchester-encoded frame where some starting part and some ending part of the Manchester-encoded signal are both colliding with other transmissions.

5. Luleå

Which of the following are true statements about the assigned reading *Small Forwarding Tables for Fast Routing Lookups* by Degermark *et al.* and associated lecture discussion?

- A. All Class A network routing entries will reside in Level 1 of the tree.
- B. The codeword array *code* is pre-computed, and does not need to be regenerated when the routing table changes.
- C. Sparse level-two chunks directly store indices into the next-hop table.
- D. Unlike search in Radix trees, the Luleå algorithm requires back-tracking up the tree data structure.
- E. In a Luleå tree with 2^{16} CIDR entries of the /16 type, the contents of the last element of the base index array contains hexadecimal `0xfff0`.

6. iSLIP

Which of the following are true statements about the iSLIP algorithm for crossbar scheduling in Internet routers?

- A. In iSLIP's grant phase, input j collects requests and grants the next request its grant pointer g_j points to.
- B. In a two-input, two-output crossbar where Input 1 always has traffic only for Output 1, and Input 2 always has traffic only for Output 2, iSLIP will schedule Input 1 to Output 1 and Input 2 to Output 2 at every iteration.
- C. In a five-input, one-output crossbar where Input 1 always has traffic for all outputs, iSLIP will schedule Input 1 to each output in a round-robin fashion.
- D. In a five-input, one-output crossbar where Input 1 always has traffic for all outputs, Input 1's accept pointer will point to "3" after the third iteration.
- E. For an N -input, N -output virtual output queued router running iSLIP realized in hardware, N bits describe the empty/non-empty state of all input queues.

7. CDMA

Which of the following are true statements about Code Division Multiple Access (CDMA) as described in lecture?

- A. If User 1 uses code c_1 and User 2 uses code c_2 , where c_1 and c_2 are chosen from a Hadamard matrix, then c_1 is orthogonal to c_1 , and c_2 is orthogonal to c_2 .
- B. To separate user i 's signal from an incoming signal where other users are transmitting simultaneously, a CDMA receiver can multiply with the inverse of user i 's code (*i.e.*, user i 's code multiplied by -1).
- C. In theory, codes chosen from an $n \times n$ Hadamard matrix can support n simultaneous CDMA users.
- D. The CDMA code rate of data encoded with an $n \times n$ Hadamard matrix is n^2 .
- E. CDMA senders listen before they send to see if the channel is occupied.

8. DNS Cache Poisoning

Which of the following statements are true about DNS cache poisoning attacks?

- A. The attacker may misdirect future queries by the victim DNS server for the target domain D to the attacker's DNS server by lying about the glue A records for the nameserver for domain D in a spoofed reply sent to the victim DNS server.
- B. Causing the victim nameserver to launch multiple queries for names in the target domain D allows poisoning to succeed more quickly.
- C. The attacker must be on the physical Internet path between the victim DNS server and the nameserver for the domain being hijacked.
- D. Setting randomly chosen query IDs on DNS queries is sufficient to make DNS cache poisoning attacks impractical.
- E. To poison the cache of a victim DNS server successfully, the attacker must win a race with a legitimate nameserver to deliver a reply to the victim DNS server's query first.

9. Isochronous vs. Asynchronous Networks

Which of the following statements are true about isochronous and asynchronous networks?

- A. Asynchronous networks waste more network capacity than isochronous ones when carrying traffic that arrives in bursts.
- B. In an asynchronous network with substantial memory for queues (enough to buffer several seconds of traffic), when link speeds are sufficiently high (*e.g.*, links running at multiple Gbits/s), one can achieve 100% link utilization with minimal queueing delay.
- C. If an isochronous network admits (*i.e.*, allows a new) connection, the end-to-end propagation delay for that connection will not change during the lifetime of the connection.
- D. No destination address is required on data sent through an isochronous network.
- E. In an isochronous network, the throughput achieved by an admitted connection depends on the traffic demands of all other connections that share links with it.

10. TCP

Which of the following statements are true about TCP?

- A. There are four packets exchanged between end hosts to establish a new TCP connection.
- B. In the wide-area Internet, TCP retransmits too aggressively when it uses a retransmit timeout (RTO) of double the mean round-trip time (RTT).
- C. TCP connections are unidirectional—they only support sending data in one direction, and the other direction is used exclusively for ACKs.
- D. If the TCP sender has not received an ACK for a data packet that it sent previously, then that data packet has not reached the receiver.
- E. As the round-trip time (RTT) on a path increases, the window size TCP must use to achieve the path's full capacity increases.

[Total for this part: 33 marks]

END OF PAPER