IIT Kharagpur
SMARTPHONE COMPUTING
and
APPLICATION (CS60009)
End-semester Examination (Fall 2016-2017)

Duration: 3 Hrs **FM: TBD**

1. Answer all questions **[5x2=10]**

   (a) What is Levy Walk? Does human walk follows Random Way Point (RWP) or Brownian Motion (BM)?
   Answer: Levy walks (LW) are characterized by a power distribution of flights where a flight is the distance that a walker travels without making a pause or a directional change.

   Human walks are statistically and fundamentally different from walks generated from commonly used mobility models such as Random Way Point (RWP), random direction and Brownian motion (BM) whose flight distributions have a short tail

   (b) What is Bursty Spot Model (BSM)?
   Answer: We generate bursty hot spots by emulating the burstiness of visit points measured from real traces. Then for a daily trip, each mobile node chooses its visit points from these points also in a bursty manner. Synthetic walk traces are generated by visiting these points using a function of 1/d a where we vary a from 1 to 3. We call our model bursty spot model (BSM).

   (c) Draw the architecture of MicroCast and describe working of each component in one sentence.

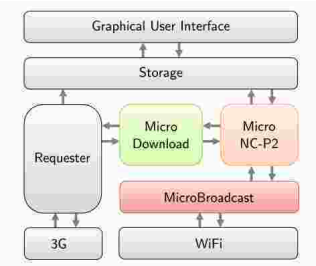

Figure 2: Architecture of MicroCast.

   - MicroDownload – Only runs on one of the phones that initiate download
   - MicroNC-P2 – Distributing segments using local wireless network
   - MicroBroadcast – Pseudo-broadcast over WiFi
   - Requester – Retrieves segments of video from the video source
   - Storage
   - Graphical User Interface(GUI)

   (d) Define Reception Rate and give the relation between cellular network rate and local network rate.
   Answer: $R_l > (N-1)R_c$
   Where R_l is Receive rate in local, R_c is Receive rate in cellular, and N is the number of phones

(e) In the context of *Tesselation*, arrange the following steps in their correct order: (i) Session block generation, (ii) Traffic attribution, (iii) Culling of traffic markers, (iv) OSN ID extraction.
Answer: The correct order is as follows: (i) OSN ID extraction, (ii) Session block generation, (iii) Culling of traffic markers, (iv) Traffic attribution.

2. (a) What are the four fundamental statistical features of human mobility? Are they independent or dependent of each other? **[4x0.5]+1** Answer:

- (F1) Truncated power-law flights and pause-times.
- (F2) Heterogeneously bounded mobility areas
- (F3) Truncated power-law inter-contact times (ICTs)
- (F4) Fractal waypoints.

All these patterns (F1F4) are intrinsically related to each other.

(b) Consider the distance matrix given below. Each entry (i,j) in the matrix represent the distance between i and j. The graph is undirected. Considering the *least-action-trip-planning* LATP algorithm, determine the order in which the vertices are visited and also the total distance covered when the start vertex 1) s=p2 and 2) s=p3. Assume the distance function to be d^{-a} where the value of a is infinite.
Explain your answer briefly. **[3+3]**

	p1	p2	p3	p4	p5	p6	p7	p8
p1	-	6	5	8	∞	∞	∞	∞
p2	6	-	4	7	5	∞	8	∞
p3	5	4	-	∞	7	6	8	∞
p4	8	7	∞	-	∞	5	∞	8
p5	∞	5	7	∞	-	5	∞	7
p6	∞	∞	6	5	5	-	∞	6
p7	∞	8	8	∞	∞	∞	-	9
p8	∞	∞	∞	8	7	6	9	-

Answer:
for S=P2
P2 –(4)–> P3 –(5)–> p1 –(8)–> P4 –(5)–> P6 –(5)–> P5 –(7)–> P8 –(9)–> P7
TOTAL 43
for S=P3
P3 –(4)–> P2 –(5)–> P5 –(5)–> P6 –(5)–> P4 –(8)–>P1
TOTAL 27
OR
P3 –(4)–> P2 –(5)–> P5 –(5)–> P6 –(5)–> P4 –(8)–> P8 –(9)–> P7
TOTAL 36

3. (a) Fill in the blanks : **[1.5]**
   i. COMBINE is based on HTTP level stripping. (TCP/HTTP / FTP)
   ii. WLAN offers much higher speeds than WWAN. (higher/lower/same)
   iii. A user who is very keen not to deplete their battery would set Ks to a larger value. (smaller/larger/zero)

(b) What are the requirements that a practical accounting scheme should ideally have ? Explain. **[3]**

- Storing Credits
- Cheat-Proof
- Privacy
- Flexibility
- Efficiency

(c) i. A node wants to download 4000 kB of data and can afford a cost of 1500 units. The system allows collaborative download and there are 5 other nodes in the system who have sent their bid and their WWAN speeds. They are specified in the table given below. The initiator node will recruit 3 collaborators. If the initiator uses threshold-based group selection criteria then which of these nodes will be selected as collaborators by the initiator. The file is divided into 8 chunks

	Bid	WWAN Speed
Node A	0.2 units/kB	20 kB/sec
Node B	0.8 units/kB	10 kB/sec
Node C	0.3 units/kB	15 kB/sec
Node D	0.2 units/kB	17 kB/sec
Node E	0.35 units/kB	8 kB/sec

of equal size. Once the 3 collaborators are selected they will download parts of the file for the initiator. If the work-queue algorithm is followed then how much cost will be incurred by the initiator for downloading the whole file. **[1.5+4]**

Answer:

Download Size = 4000 kB

Cost = 1500 units

TC = 1500 / 4000 = 0.375 unit/kB

So the collaborators who qualify are : Node A, Node C, Node D, Node E. Based on the WWAN speed the collaborators selected are : Node A, Node D, Node C.

The file is divided into 8 chunks of equal size. Each chunk size is 500kB.

Node A	Node D	Node C	Node A	Node A	Node D	Node C	Node A
1	2	3	4	5	6	7	8
25 sec	29.41 sec	33.33 sec	50 sec	58.82 sec	66.67 sec	75 sec	88.23 sec

Node A takes 0.2*500 = 100 units.

Node D takes 0.2*500 = 100 units

Node C takes 0.3*500 = 150 units

Total Cost = (100*3) + (100*3) + (150 *2)=900 units

ii. What are the two basic group selection criteria ? Which one of these group selection criteria can be posed as an optimization problem and how ? Briefly Explain what the terms stand for ? Also, for the conservation group selection criteria explain how initiator calculates the total cost ? **[1+3+1]**

- **Threshold Based Group Selection - Conservative**
  Consider an initiator willing to download a file of size F and is willing to incur a cost of C then the TC = C/F
- **Opportunistic Group Selection - Optimization**

Number of collaborators = N
The goal of the optimization is to minimize the total time taken to download F bytes of data subject to the cost constraint C. If each collaborator i, working in parallel, downloads xi bytes with bandwidth Bi, the total time taken to download the file is the maximum of $\{x_1/B_1, x_2/B_2, \ldots, x_N/B_N\}$, where N is the number of distinct *I-am-Alive* packets received by the initiator. We determine optimum values of x_i, $i = 1 \ldots N$ so that we minimize the total time subject to the constraints:

$$\sum_{i=1}^{N} TC_i X_i \leq C$$

$$\sum_{i=1}^{N} xi_i = F$$

$$x_i \geq 0, i = 1 \ldots N$$

4. (a) Consider that a network of 4 phones are there. Each phone has a download speed and download cost as shown below: Consider that a file of 240 kb is to be viewed in

Phone 1	5 kbps	Rs 1/kb
Phone 2	15 kbps	Rs 2/kb
Phone 3	30 kbps	Rs 5/kb
Phone 4	30 kbps	Rs 5/kb

this network and segment size is 30 kb. There is no backlog in any device and there can be maximum backlog of 2 (value of K). Once a download in a device is started, it is not cancelled by the system and rescheduling of download task not started is done, only in case any of the device is free.

i. Use a timeline of download task per second for each phone and explain your answer. **[3]**

ii. Give out the time and cost incurred in downloading the file in a MicroCast environment. **[2]**

(a) File size = 240 kb ; Segment size = 30 kb

No of segments = 240/30 = 8 segments

	Initial segment Allotment	At time 1 sec	At time 2 sec	
Phone 1	[1] [2]	5 kb of segment [1]	10 kb segment [1]	
Phone 2	[3] [4]	15 kb of segment [3]	Complete segment [3]	
Phone 3	[5] [6]	Complete segment [5]	Complete segment [6] and Idle	Segment [2] of phone 1 reschedule to Phone 3
Phone 4	[7] [8]	Complete segment [7]	Complete segment [8] and Idle	

cont-d

	At time 3 sec	At time 4 sec	At time 5 sec	At time 6 sec
Phone 1	15 kb of segment [1]	20 kb of segment [1]	25 kb of segment [1]	Complete segment [1]
Phone 2	15 kb of segment [4]	Complete segment [4]	Idle	Idle
Phone 3	Complete segment [2]	Idle	Idle	Idle
Phone 4	Idle	Idle	Idle	Idle

(b) Time required will be 6 seconds

Cost based on download by each phone

Phone 1 → 1 segment → 30 kb = Rs 30

Phone 2 → 2 segments → 60 kb = Rs 120

Phone 3 → 3 segments → 90 kb = Rs 450

Phone 4 → 2 segments → 60 kb = Rs 300

Total cost = Rs 30 + 120 + 450 + 300

= Rs 900

(b) Explain the steps in encoding and decoding used in MicroCast. **[1+1]**

Answer

Encoding:

- One segment is divided into m packets each of size n. Let B matrix represents all the m packets.
- Size of B is mxn.
- Let's take identity matrix I of size mxm
- Augment I with B ie [B|I]
- Randomly generate coefficient matrix R of size mxm
- Encoding R*[B|I]

The encoded message is sent to the network. Let R^{-1} = Inverse of R

Decoding:

R^{-1} x Encoded Matrix
= R^{-1} x (R x [B|I])
= (R^{-1} x R) x [B|I]
=I x [B|I] = [B|I]

5. This question is based on the TailEnder system. The question has 2 parts:

(a) Consider the TailEnder scheduling algorithm. What are the conditions under which a request is enqueued instead of being transmitted immediately? Please explain the meaning of any notation that you use in your answer. **[3]**

Answer: The following conditions need to be met: (1) $t < d_i$; (2) $(d' + 0.57T) \geq a_i$. t is the current time, d_i is deadline, d' is the previous deadline, and T is the *tail time*.

(b) A user queries for "Android app development"on Bing. The first 5 search results appear in the following order: (1) developer.android.com, (2) codeproject.com, (3) tutorialspoint.com, (4) wikipedia.org, (5) vogella.com. Assume that the user tosses a fair coin 5 times consecutively, and visits the first k pages (in order of their ranks) where k is the number of heads. Assume that tail energy is 40 units, and the energy required to fetch k pages is given by $(k/2)$ units. If the total energy required to receive a document is 60 units (including ramp-up energy, tail energy, and transfer energy required to receive a requested document), calculate the expected energy savings in the following cases: **[2x3=6]**

i. The first 2 pages (in order of rank) are prefetched.
ii. The first 4 pages (in order of rank) are prefetched.
iii. All 5 pages are prefetched, but the tail energy is now 112 units and the total energy requirement increases to 120 units.

Answer: Let X be a random variable which denotes the number of heads in 5 consecutive coin tosses. Therefore, $P(X = k) = \binom{5}{k}(\frac{1}{2})^{5-k}(\frac{1}{2})^k = \binom{5}{k}(\frac{1}{2})^5$. Therefore, probability that the user visits 2, 4, and 5 pages are $\frac{10}{32}$, $\frac{5}{32}$, and $\frac{1}{32}$, respectively.
Tail energy = 40 units.
Energy required to fetch 2, 4, and 5 pages are 1, 2, and $\frac{5}{2}$ units respectively.
Total energy required to receive a document = 60 units.
(i) Expected energy savings if 2 pages are prefetched = $\frac{(\frac{10}{32}.40)-1}{60} = \frac{23}{120} = 0.19$ units
(ii) Expected energy savings if 4 pages are prefetched = $\frac{(\frac{5}{32}.40)-2}{60} = \frac{17}{240} = 0.07$ units

(iii) Expected energy savings if 5 pages are prefetched $= \frac{(\frac{1}{32}.112)-\frac{5}{2}}{120} = \frac{1}{120} = 0.008$ units.

6. Consider the Bartendr system for energy-aware cellular data scheduling. The dynamic programming algorithm for computation of the minimum energy schedule is as follows:

```
Initialization:
for t = 1 to M do
   E_{0,t} = 0
end for

Computing optimal schedules:
for k = 1 to N do
   for t = k to M do
      E_{k,t} = Compute_E_{k,t}()
      Last_{k,t} = l value for which the previous quantity was minimized
   end for
end for
```

In the above algorithm, $E_{k,t}$ is the minimum energy required to transmit k frames in t time slots, and $Last_{k,t}$ stores the slot number where the k^{th} frame is scheduled.
Explain how the function **Compute**_$E_{k,t}()$ is implemented. Explain the meaning of any new notation that you use. Also, comment on the complexity of the algorithm. **[3+2]**
Answer: The implementation is as follows:
(1) Compute_$E_{k,t}()$:
(2) $E_{k,t} = \min_{l=k-1}^{t-1}(E_{k-1,l} + ESlot_{l+1})$
(3) Return $E_{k,t}$
$ESlot_{l+1}$ is the sum of the communication energy required to transmit a frame in slot $(l+1)$ and the incremental tail energy cost, given the transmissions that occurred in the previous slots.

The complexity of the dynamic program is $O(M^2.N)$, where N frames are transferred in time M.

7. Answer the following questions in the context of *Tesselation*:

(a) The following is a table with details of 3 session blocks: **[2.5+2.5]**

Session Block	Traffic-marker Pairs	Duration
S1	(u_1,m_1), (u_2,m_2)	5 units
S2	(u_1,m_3), (u_2,m_1)	10 units
S3	(u_3,m_1)	15 units

i. Calculate the *uniqueness* of each of the traffic-marker pairs seen in the table.
Answer: $P(u_1, m_1) = \frac{5}{30} = \frac{1}{6}$; $P(u_2, m_2) = \frac{5}{30} = \frac{1}{6}$;
$P(u_1, m_3) = \frac{10}{30} = \frac{1}{3}$; $P(u_2, m_1) = \frac{10}{30} = \frac{1}{3}$; $P(u_3, m_1) = \frac{15}{30} = \frac{1}{2}$
Uniqueness of $(u_1, m_1) = 1 - \{P(u_2, m_1) + P(u_3, m_1)\} = 1 - \{\frac{1}{3} + \frac{1}{2}\} = \frac{1}{6}$
Uniqueness of $(u_2, m_2) = 1 - 0 = 1$
Uniqueness of $(u_1, m_3) = 1 - 0 = 1$

Uniqueness of $(u_2, m_1) = 1 - \{P(u_1, m_1) + P(u_3, m_1)\} = 1 - \{\frac{1}{6} + \frac{1}{2}\} = \frac{1}{3}$
Uniqueness of $(u_3, m_1) = 1 - \{P(u_1, m_1) + P(u_2, m_1)\} = 1 - \{\frac{1}{6} + \frac{1}{3}\} = \frac{1}{2}$

ii. Calculate their *persistency* as well, if possible. If not possible, justify.
Answer: *Persistency* is defined only for pairs for which *uniqueness* $= 1$. Therefore, *persistency* is defined only for the pairs (u_2, m_2), and (u_1, m_3).
$P(u_1) = P(u_1, m_1) + P(u_1, m_2) + P(u_1, m_3) = \frac{1}{6} + 0 + \frac{1}{3} = \frac{1}{2}$
$P(u_2) = P(u_2, m_1) + P(u_2, m_2) + P(u_2, m_3) = \frac{1}{3} + \frac{1}{6} + 0 = \frac{1}{2}$
Persistency of $(u_2, m_2) = 1 - \frac{P(u_2,m_1)+P(u_2,m_3)}{P(u_2)} = 1 - \frac{\frac{1}{3}+0}{\frac{1}{2}} = \frac{1}{3}$
Persistency of $(u_1, m_3) = 1 - \frac{P(u_1,m_1)+P(u_1,m_2)}{P(u_1)} = 1 - \frac{\frac{1}{6}+0}{\frac{1}{2}} = \frac{2}{3}$

(b) State the criterion for selecting the *activity fingerprint* F_i for user u_i. **[1]**
Answer: The top k most frequently accessed services are selected as the *activity fingerprint*.

(c) The activity fingerprint of a user u_1 is given by $F_{u_1} = \{a_1, a_2, a_3\}$. The list of services used by 4 other users u_2, u_3, u_4, and u_5, are given by:
$S_{u_2} = \{a_1, a_2, a_4, a_5\}$, $S_{u_3} = \{a_1, a_2, a_3, a_5\}$, $S_{u_4} = \{a_1, a_3\}$, $S_{u_5} = \{a_1, a_2, a_3, a_4\}$
Compute the uniqueness of fingerprint F. **[3]**
Answer: Here, $k = |F_{u_1}| = 3$. Among services for users u_2, u_3, u_4 and u_5, $|S_{u_2}| = |S_{u_3}| = |S_{u_5}| \geq k$. Therefore, S^k = {set of users having at least k activities} = $\{u_2, u_3, u_5\}$, which means $|S^k| = 3$. Now, $F^k_{u_1}$ = {set of users having F_{u_1} as a subset of their activities} = $\{u_3, u_5\}$, which means $|F^k_{u_1}| = 2$.
Uniquess of $F_{u_1} = 1 - \frac{|F^k_{u_1}|}{|S^k|} = 1 - \frac{2}{3} = \frac{1}{3}$.

8. Answer the following questions in the context of *OverLay* – a system for practical augmented reality.

(a) What do you mean by two objects A and B being *spatially invariant*? Why is conditional spatial invariance called *conditional*? **[2+1]**
Answer: Objects A and B are spatially invariant when A and B are coplanar with a room's wall (e.g., mounted). *OverLay* assumes that A and B could be spatially invariant, if they have been found simultaneously in view by one or more observers.
In case of *conditional spatial invariance*, the condition does not hold universally. The invariance holds only in a specific region.

(b) What clustering algorithm is used for preparing the secondary matching database during learning from retrieval? What is the advantage and disadvantage of using this secondary database? **[1+2]**
Answer: The *KMeans* clustering algorithm is used.
Advantage: When the primary database is insufficient, the secondary database provides valuable matches.
Disadvantage: The secondary database introduces a slight delay to the response.

(c) OverLay uses linear programs to minimize errors in which two kinds of positions? **[2]**

Answer: The positions are: (1) Relative angular (rotational) position, (2) Relative temporal spacing/position.

(d) *"We consider the rotational distance from any tag i as $|P_i - P_U|$ mod $2\pi + E_*^R(i)/2$. We include only half of the error term to ..."*. Why is only half of the error term included? **[1]**
Answer: Only half the error term is included to reflect the $50-50$ probability that the error should work in favour or against accepting tag i.

9. Answer the following questions in the context of *Glimpse* – a system for continuous, real-Time object recognition on mobile devices.

(a) Mention the factors which dictate the selection of parameter p in the following equation: $p = l/n$, where n is the total number of frames in a duration, and l out of them are to be selected from the active cache. **[2]**
Answer: The selection of p depends on 2 factors: (i) the end-to-end processing delay of a frame (i.e., the value of n), and (ii) the execution time e of the tracking algorithm on the client.

(b) Suppose that the absolute difference for pixel values (x, y) between frames i and j in grayscale is given by $a_{i,j}(x, y)$. How do you compute the frame difference metric $d_{i,j}$? What is the complexity of this computation? **[2+1]**
Answer: The difference metric for pixel (x, y) is given by:

$$d_{i,j}(x,y) = \begin{cases} 1 & a_{i,j}(x,y) > \phi \\ 0 & otherwise \end{cases}$$

The frame difference (movement metric) $d_{i,j}$ between frame i and frame j is given by:

$$d_{i,j} = \sum_{x,y} d_{i,j}(x,y),\ d_{i,j} \geq 0$$

Complexity: Computing $d_{i,j}$ is linear in the size of the frame.

(c) The frame selection problem can be mapped to an existing problem – name this problem. Can you comment on the complexity of the dynamic program that solves this problem? **[1+1]**
Answer: Linear partition problem.
Complexity: The dynamic problem can be solved in $O(ln^2)$ time, where n is the number of frame differences, and $(l+1)$ is the number of partitions.

(d) Suppose that $H[n, l]$ is the optimum value of a partition arrangement with n frame differences and l partitions. Please write the recursive formulation for computing $H[n, l]$. **[3]**
Answer: The formulation is given as:

$$H[n,l] = \min_{j=i}^{i+n}(\max\{H[j,l-1], \sum_{k=j}^{i+n} d_{k,k+1}\})$$

End