Honor Code of the School of Engineering

“All students taking courses in the School of Engineering agree, individually and collectively, that they will not give or receive unpermitted aid in examinations or other course work that is to be used by the instructor as the basis of grading.”
-From the Graduate/Undergraduate Bulletin

I have read, understood, and agree to abide by the Honor Code of the School of Engineering.

Name: ____________________________________________

ID: ____________________________________________

Signature: ____________________________________________

Date: ____________________________________________

1. [10 points] Please give the whole names for the following acronyms: HTC, VPC, CUDA, CPS, DoP, EBS, HIDS, ILP, MPP, MTTR.

2. [40 points] True or false (yes or no, 1 or 0) problems with wrong-answer penalties:
   a) Compute and storage services belong to PaaS.
   b) The client-server approach has excellent scalability.
   c) Fabric computing prefers to use high speed bus to connect all servers in the data center.
   d) GPU can do graphic and DSP operations, but cannot do general purpose application.
   e) Parallel data transfer uses two or more data streams over the same channel, striped data transfer uses two or more streams in different channels to access different blocks of the file for simultaneous transfer.
   f) A CPU architecture is virtualizable if it supports the ability to run the VM’s privileged and unprivileged instructions in the CPU’s user mode while the VMM runs in supervisor mode.
   g) Napster has a constant search time.
   h) The most important criteria for wireless sensor is cost-effectiveness because we need to deploy lots of sensors randomly but still make them connected under the area of interest.

3. [20 points] Which of the following are not for multi-tenancy? a) virtualization, b) centralization of infrastructures in locations, c) deployment of multiple clients in a single software instance, d) improve utilization, e) improve efficiency, f) increase peak-load capacity.

4. [50 points] Consider a program for multiply two large-scale $N \times N$ matrices, where $N$ is the matrix size. The sequential multiply time on a
single server is $T_1 = cN^3$ minutes, where $c$ is a constant determined by the server used. An MPI-code parallel program requires $T_n = cN^3/n + dN^2/n^{1.5}$ minutes to complete execution on an $n$-server cluster system, where $d$ is a constant determined by the MPI version used. Assume the program has a zero sequential bottleneck ($a = 0$). The second term is $T_n$ accounts for the total message-passing overhead experienced by $n$ servers. Answer the following questions for a given cluster configuration with $n = 64$ servers, $c = 0.8$, and $d = 0.1$. Parts a) and b) have a fixed workload corresponding to the matrix size $N = 15,000$. Parts c) and d) have a scaled workload associated with an enlarged matrix size $N' = n^{1/3}N = 64^{1/3} 	imes 15,000 = 4 \times 15,000 = 60,000$. Assume the same cluster configuration to process both workloads. Thus, the system parameters $n$, $c$, and $d$ stay unchanged. Running the scaled workload, the overhead also increases with the enlarged matrix size $N'$.

a) Using Amdahl’s law, calculate the speedup of the $n$-server cluster over a single server.
b) What is the efficiency of the cluster system used in part a)?
c) Calculate the speedup in executing the scaled workload for an enlarged $N' \times N'$ matrix on the same cluster configuration using Gustafson’s law.
d) Calculate the efficiency of running the scaled workload in part c) on the 64-processor cluster.
e) Compare the above speedup and efficiency results and comment on their implications.

5. [20 points] In SMP, which components are single-points-of-failure?

6. [20 points] A BCube$_k$ ($k \geq 1$, with level 0 to level $k$) is built recursively from $n$ BCube$_{k-1}$ with $n^k$ $n$-port switches, each node is represented by a $(k+1)$-number (with each number from 0 to $n$-1). In case of $k = 3$ and $n = 4$, what is the length of the shortest path (in terms of hops) from node 0000 to node 3333 in the BCube$_k$ (assume one hop per switch)? How many possible shortest paths are there? How many possible parallel shortest paths are there? Please show one of the possible shortest paths (or a sequence of nodes)?

7. [20 points] Multiple choice questions with multiple answers for the questions: which of the following have consistency/coherent problem:

- a) replicates in distributed file systems
- b) L1 cache
- c) L2 cache
- d) L3 cache
- e) local memory
- f) shared memory

8. [20 points] Which of the security features in the table below are the cloud user’s or the cloud provider’s responsibility?

<table>
<thead>
<tr>
<th></th>
<th>confidentiality</th>
<th>integrity</th>
<th>availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>SaaS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PaaS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IaaS</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>