# Principles of Computer Systems
# Final Exam

19-Dec-2019

This exam has 5 questions, totaling 100 points. You have 105 minutes to answer them, which means you earn about 1 point per minute of work – please consider spending on each question no more minutes than the number of points attributed to it.

The questions are all related to one system, described on the first interior page of the exam. We recommend that you read this description and then read all the questions before you start answering any one of them. This will give you context. Please make sure you answer the question we actually ask, not something else, because a correct answer to the wrong question does not earn any points.

If you exit the room during the exam, you will have to turn in your exam, and you will not be permitted to return to the room until the end of the exam. Please plan accordingly.

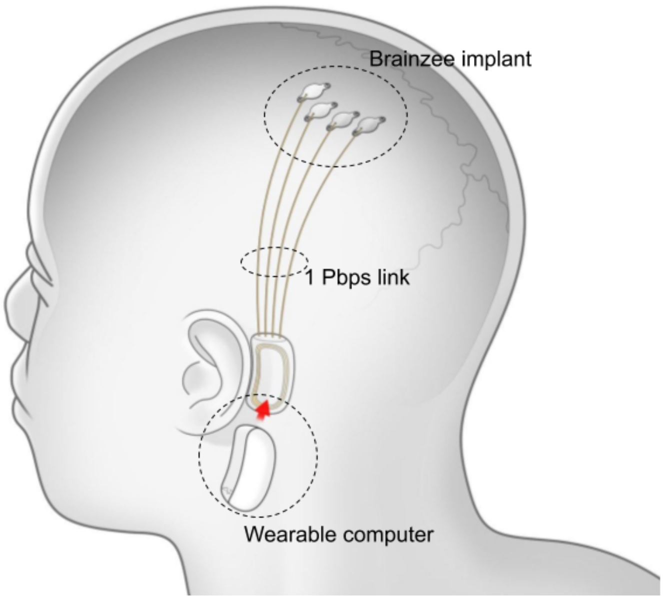
You are allowed to have any amount of printed material you like (books, papers, notes) but no laptops, tablets, cellphones, etc. are permitted during the exam. You must take the seat assigned by the course staff and present your CAMIPRO card to the staff upon request.

If you need scratch paper, you will find several sheets at the end of the exam. Writing neatly and clearly on the exam itself will greatly improve the state of mind of the grading staff.

| |
|---|
| **Do not open the exam until instructed to do so.** |

**Your name: ___________________________ SCIPER: ___________**

Imagine a future in which there exists a system, called the Collective, whose purpose is to enable every human being to have access to the knowledge of every other human being on Earth. Every participant in the Collective has a Brainzee device implanted in the brain, which essentially acts as a high-bandwidth interface between the analog signals in neurons and a digital computer that is worn behind the ear.




The wearable computer can retrieve (via the Brainzee) knowledge from the brain, and can transmit (via the Brainzee) knowledge to the brain. The computer can communicate with other Collective nodes (i.e., computers worn by other participants) via WiFi, but with performance that is on-par with being plugged into a modern Ethernet jack at all times.

Once a Brainzee is implanted, it cannot be removed. The Brainzee manufacturer has a method for giving each of its devices a globally unique identifier (which can be read by the computer) but does not have a directory of these identifiers.

In this imagined future, there exists a universal encoding standard for brain-resident knowledge (Uniknode) used for exporting / importing snippets of knowledge via the Brainzee. These snippets of knowledge can be in textual format, images, videos, smells, etc., and the Brainzee can turn them into a binary representation based on Uniknode.

The Brainzee device provides the wearable computer with a keyword/topic-based search interface: the computer can submit a query in freeform text and receive with virtually-zero latency a set of Uniknode-encoded knowledge snippets representing all the acquired knowledge in that brain that is relevant to the given topic/keywords. These snippets can be understood by other Brainzee devices, which can decode them and supply them to their host brains.

Each participant in the Collective has a profile, and the profile consists of a list of keywords, chosen by the participant, which describe in decreasing order of importance how the participant thinks of themselves (e.g., "<conservative, male, Norwegian, business executive, gay, millionaire").

You can think of the Collective as the Web on steroids. One major benefit is the removal of slow human interface devices (keyboards, screens, microphones, etc.) that currently form a bottleneck in the sharing of knowledge. Instead, we have high-bandwidth Brainzees.

Your task is to design the Collective.

## **Question 1** *(10 points)*

The point of the Collective is to enhance humans by enabling them to ask the Collective (i.e., the aggregation of all participant brains) for knowledge. The way this works is that the participant thinks of what they want to find out, and the Brainzee encodes this demand in a query to the local computer, which then takes the necessary steps to answer that query by reaching out to the Collective. For example, a participant may think "Hmm, I wonder what the Welsh think about Brexit", and the Brainzee implant turns this thought into a query that is submitted to the participant's wearable computer, and shortly thereafter the sought-after knowledge starts pouring into the requesting brain.

Design the query interface offered by the local computer to the Brainzee: How do queries look like? In what form are the results returned? Are there any reserved keywords for the query? There are many choices to be made here, and they are all left entirely up to you[1]. State your assumptions. Aim for simplicity.

[1] If you feel you absolutely need to know more about Uniknode, you can assume that it consists of a bunch of computer-readable metadata followed by a binary object that can only be read by a Brainzee implant (which will then decode it and instantiate the encoded knowledge in the corresponding brain).

## **Question 2** *(20 points)*

Think of and describe a rough-cut design of the Collective. Don't worry yet about security, trust, or privacy. You can assume that all the hardware is tamper-proof, and everybody is honest. Feel free to make any assumption you want about the wearable computer's processing power and memory capacity, just make sure you state your assumptions. We encourage you to use diagrams to explain your design.

Some things to think about:

- Does the Collective have a centralized or decentralized architecture?
- What does it do with frequently accessed knowledge?
- Can you request knowledge from specific groups or individuals?
- Does it deduplicate knowledge snippets (e.g., "World War II ended in 1945" would be known by many participants, so probably there is little use to providing millions of identical answers to a query about the end of WWII)?
- How does it tolerate failures and death?

Please avoid philosophical questions that are not related to POCS.

## Question 3 *(35 points)*

In order to answer the following four sub-questions, you will need to do some back-of-the-envelope calculations. For your convenience, we included the "systems constants" slides from lecture at the end of this exam. While there is no good estimate of how much knowledge a brain can store, the best estimate of how much *information* it can hold is 2.5 Petabytes[2] – you can use this number if you need to estimate brain capacity. There are currently 7.7 billion humans on Earth[3], and you can assume that they all participate in the Collective.

*Question 3.1*: In your design, what is the expected rate of queries received from the Collective by the computer of an arbitrary participant? How many queries-per-second would you expect the computer of a participant like Stephen Hawking or Beyoncé to receive at peak load? Could the recipient's equipment answer the queries at this rate? What if the world's population doubles? Remember to always state your assumptions.

[2] Paul Reber, "What Is the Memory Capacity of the Human Brain?", Scientific American, Neuroscience section, May 1, 2010

[3] United Nations Department of Economic and Social Affairs, "Population Dynamics: World Population Prospects", October 2019

*Question 3.2*: What is the maximum latency with which the Collective (as designed by you) may answer a query? In other words, what is the longest that a Brainzee would have to wait to receive an answer to a query it submitted to its attached computer (which then may have gone on to ask the Collective)? Justify quantitatively.

*Question 3.3*: What is the minimum latency with which the Collective (as designed by you) would answer a query? In other words, what is the soonest that a Brainzee could expect to get an answer to a query it submitted to its attached computer (which then may have gone on to ask the Collective)? Give an example of a Brainzee-to-computer query that would incur that minimum latency, and argue quantitatively that it is indeed the minimum.

*Question 3.4*: How fresh is a knowledge snippet returned by the Collective (as designed by you) relative to the providing participant's knowledge? In other words, what is the upper bound on the amount of time that elapses between moment $t_0$ when the snippet is provided by the originator's Brainzee and moment $t_1$ when the recipient's Brainzee receives it? Describe your reasoning quantitatively.

## Question 4 *(15 points)*

Could a denial-of-service attack be perpetrated in the Collective with the purpose of preventing a particular participant (e.g., a political dissident, a contrarian scientist) from sharing their knowledge? If yes, provide an attack scenario and quantify the maximum scale it can achieve (you can choose how to describe scale). If no, quantitatively explain why not.

*Note*: Your design does not have to be immune to denial-of-service attacks to receive full credit for Question 2. The point of this Question 4 is for you to show how you reason about denial of service.

## **Question 5** *(20 points)*

The way the Brainzee implant works is that it directly extracts knowledge from the brain, therefore one cannot purposely make up knowledge or information and pretend to the Brainzee that the belief in that knowledge is genuine – the Brainzee would witness the process of making up the fake knowledge.

This implies that the Collective can only propagate true beliefs. However, what one truly believes could be true ("WWII ended in 1945") or false ("WWII ended in 1944").

Therefore, participants would like to know who originated a particular knowledge snippet, in order to decide whether to count on it (e.g., if you wonder "are vaccines safe?" you might want to know what scientists think vs. non-scientists). Of course, disclosing the Brainzee ID with each snippet would pose serious privacy concerns. Participants however may be willing to disclose their identity for certain classes of snippets (e.g., "It's OK for people to know I'm Stephen Hawking for any topics related to time and black holes").

How would you modify your system to accommodate the ability of knowing the origin of knowledge snippets? What guarantees would your design provide to the recipients of knowledge and to the originators of knowledge?

*Slides 29-30 of the 17-Oct-2019 lecture are reproduced here for your convenience*

## Systems “Constants” (CPU cycles)

- Register-register ADD/OR/etc. < 1 CPU cycle
- Memory write ~1 cycle
- Correctly / incorrectly predicted "if" branch = 1-2 cycles / 10-20 cycles
- L1 / L2 / L3 / main RAM read = 3-4 cycles / 10-12 cycles / 30-80 cycles / 100-150 cycles
- TLB hit / miss = 0.5 - 1 cycle / 7-21 cycles
- CAS = 15-30 cycles
- C function direct / indirect call = 15-30 cycles / 20-50 cycles
- Kernel syscall = 1,000 - 1,500 cycles
- Thread context switch (direct costs) = 2,000 cycles
- On NUMA, different-socket mem hierarchy access is 3 - 10x that of non-NUMA

## Systems “Constants" (absolute time)

- L1 / L2 / main RAM reference = 1 / 4 / 100 ns
- Mutex lock or unlock = 17 ns
- Branch misprediction = 3 ns
- Send 2KB (2,000 bytes) over commodity network = 88 ns
- Compress 1 KB with Snappy = 2,000 ns = 2 microsec
- SSD random read = 16,000 ns = 16 microsec
- Read 1 MB sequentially from RAM = 5,000 ns = 5 microsec
- Packet roundtrip in same datacenter = 500 microsec = 0.5 millisec
- Read 1 MB sequentially from SSD = 78 microsec
- Seek on magnetic disk = 3 millisec
- Read 1 MB sequentially from magnetic disk = 1 millisec
- Packet roundtrip CA -> Amsterdam -> CA = 150 millisec