T-101.5241 Exam

Eduardo Castellanos (397182)

eduardo.castellanos@aalto.fi

December 15, 2014

Question 1

Give a taxonomy of peer-to-peer (P2P) networks.

Compare the advantages and disadvantages of each category.

The following is the taxonomy for the P2P networks as given by Prof. Jukka Nurminen[1]. They can be broadly split into structured, and unstructured.

Centralized. This was the first generation. In the centralized scheme there is a single server that coordinates all the peers. All the indexing data is stored on the server, but the content is only available from the peers. **Advantages** Searches are faster and more complete. The architecture is simple. Advertising can be used to monetize it. **Disadvantages** Single point of failure. The central server can be targeted (legally or by a DoS attack) to bring the whole network down. Upkeep of the central server becomes expensive. Central server can be a bottleneck.

Pure. In a pure peer to peer system, all the parties involved function as servers and clients. Peers establish connections between themselves randomly. Every peer keeps a set of connections to other peers and performs searches by query flooding. **Advantages** The system is fully distributed, and resilient. There is no single point of failure. Can provide anonymity. **Disadvantages** Query flooding is an expensive and inefficient. The results are often not comprehensive. There is no responsible party, difficult to monetize. The overlay topology is not optimal, it can cause load inequality, loops, and zigzag routes.

Hybrid. The main differentiation with Pure P2P is the incorporation of an additional hierarchical layer. There is an election process to vote the superpeers. **Advantages** Reduces the load on the network since querying is more controlled. Reliable and resilient since there's no single point of failure. Can provide anonymity. **Disadvantages** Still high signalling traffic, and the search is not 100% effective. There's more load placed on the superpeers. The overlay topology still suffers from the same ailments as in Pure

P2P (although to a lesser degree).

DHT-Based. Distributed Hash Tables are used to index the content, and they are dispersed among the peers. **Advantages** It is robust, and the search is comprehensive. Highly scalable. **Disadvantages** No Fuzzy queries. The network cannot quickly adapt to structure changes because it needs to rebuild the tables.

Strengths and weaknessess of P2P technology.

Describe the strengths and weaknesses of P2P technology as a whole (from both technical and non-technical perspectives).

Strengths. It is extremely scalable since it's all about sharing resources. This makes it resistant to flash crowds. It significantly decreases the load on content servers. Therefore, it reduces the need of infrastructure, which brings the costs down. Anyone can bring his own content. In addition, it provides high availability since the content is distributed across the peers.

Weaknesses. P2P networks can be complex and dificult to police and manage. They can be easily abused to distribute illegal or malicious content.

Question 2

Describe the approaches to improve the energy efficiency of mobile platforms. Explain in detail.

Mobile platforms are complex systems, therefore there is no one single component that has to be targeted for energy efficiency. However there are several approaches that have been persued to achieve better effciency.

Do less work. Do not implement all the functionality of network protocols or applications in mobiles, but restrict it to the essential. In P2P networks acting strictly as a client instead of serving content as well improves energy, but it also affects the tit-for-tat mechanism.

Proxy assisted downloading. Many tasks the mobile needs to perform could instead be performed in the cloud, thus alleviating the need for processing. Hosting a proxy in the cloud that shapes traffic to make it more energy efficient can also aid in improving energy efficiency.

Transfer in scheduled bursts. Redesigning applications to transfer data in high bursts brings significant energy savings because the power to transmit a bit decreases as the transmission bit rate increases. An example is BurstTorrent.

Delayed data transfer. Transferring data when performing a call is more energy efficient than doing both separately. Traffic can be synchronized to fill spare capacity with parallel connections. This is extremely effective with VoIP calls.

Optimize hardware. The integrated circuits can be made to be more energy efficient. Also, the CPUs are underclocked to reduce the energy consumption, but still remain powerful enough for use. Not only ICs can be improved, but also radio, sensors, input methods, etc. The design of more energy efficient electrical components plays an important role in energy efficiency.

Network communication. Another approach is to optimize network communication. It has been shown that sending data in bursts is more efficient than keeping a continuous and slow connection. This takes advantage of the power saving mechanisms in radio equipment. This approach decreases the tail energy lost. Designing energy efficiency aware communication protocols is another way of improving the efficiency. Eliminating the need for keep-alive messages in connections can impact power consumption by a factor of 10 or more. Unwanted traffic also consumes energy: port scans, malware, misconfigured devices, etc.

Question 3

Describe the techniques to distribute content (including streaming and non-streaming services).

Streaming

Server This is the most widely used because of it's simplicity and ease of set up. It involves setting up a regular server which can be accessed by client's browsers. Expensive to scale, limited quality and capacity.

Server grid Uses a content delivery network which balances the load of the connections between the different servers. Still expensive to scale.

IP multicast / LAN multicast This is the ideal model, but due to technical constraints and lack of motivation on the provider's end to configure and implement it, it has hot been deployed widely. It can work in local environments where the infrastructure is controlled. Relies on the IP Multicast IETF standards.

Peercasting (P2P streaming) Each user that is receiving the stream will also forward it to other users. In theory, this would not require servers, and have an "infinite" scalability. However, there are many challenges with this model because of the churn, asymmetric data connections, limited peer visibility because of firewalls and NATs, and using the resources optimally.

Multicast tree It was the first approach with over 20 variants where the peers form a tree topology for each data stream. It works in practice, but it requires large output bandwidth. Also tree optimization and repair are not straight forward, and in practice less than half of the peers can contribute.

Data-driven overlay This is not the best, but the most widely used approach. There is a lot of active research in the area. It is analogous to BitTorrent in that it splits the stream in small pieces and distributes them to the swarm. Most large-scale solutions use it because it scales well. The topology is created using a gossiping protocol. Peers maintain a random partial view of the network from selected random partners because there is no centralized tracker. There are still many challenges that are being researched. The main problems are flash crowds and dealing with 50% of users that don't contribute.

Hybrid technologies This is the best known technology for commercial large-scale streaming. It solves the problems of the data-driven overlay by combining P2P and a server grid. In stable conditions a P2P distribution is used, and then powerful servers are used to complement the required capacity. The servers also support newcomers, and aid users behind asymmetric connections.

Non-streaming

Web proxy caches Web caches can be implemented in different locations (Reverse vs. Forward). It works by retaining a copy of the content in the cache, so when a client requests a resource it can serve it more quickly. There are limitations to the capacity since the objects in the cache can be of different sizes, and there is only so much that a cache can hold. Eviction strategies must be implemented in order to free up the cache.

Cooperative web caches Works by sharing the cache between a number of cache servers. This approach works best for relatively small populations of users, for larger populations it offers no significant advantages[2].

Content Delivery Networks Content delivery networks have servers located in several places around the world, and provide means to deliver the content to the users from the server that is closest to them.

Push CDN In a Push CDN, every edge server has an up-to-date replica of the content before it is even requested.

Origin pull CDN In an origin pull CDN, the edge server gets a copy of the content the first timeit performs a request, and subsequently serves the content from the cache.

Application CDN Replicates the whole server side application in geographically disperse servers.

Peer-to-Peer Aids in the distribution of content, by having the peers send the chunks of data among themselves, instead of using up all the bandwidth of the server.

Cache + CDN It has been studied, but proven to be ineffective[3].

Question 4

Nano communications

Explain the principles of nano communications.

In nano communications one communication approach is with electromagnetic nano communications. Graphene is one of the most promising materials for this task, it can be used as an antenna. Energy harvesting can be done by using zinc oxide nanowires. Communication could be done in the terahertz band. However, there are sill challenges regarding the battery power, line of sight requirement, and the molecular interaction/absorption.

Another approach is using molecular communication. This approach relies on encoding information into organic molecules such as DNA, proteins or peptide chains. Then these molecules are transmitted through diffusion or active transport to other nanomachines or cells. The challenges is that it is very slow and unreliable, and Directionality is not easy. This makes traditional communication approaches not suitable.

Internet of Nano Things

Describe the typical application scenarios of Internet of Nano Things.

Healthcare. Using nanosensors to to provide realtime data on body status and functions. They could also be used to analyze body fluids instead of doing lab tests. These sensors could also be placed in hospital or other public places to track viruses or other pathogens[4].

Environmental. They also could be used to monitor pollution, greenhouse gases, and radiation[4].

Agricultural monitoring. The nanosensors could also be used to detect plant or animal pathogens such as E.coli and mad cow disease[4].

Question 5

Indoor localization

Describe the indoor localization technologies, and compare their strengths and weaknesses.

Several indoor location solutions have been proposed. Researchers have developed systems for indoor localization based on many technologies such as FM radio [5], Bluetooth[6], GSM[7, 8, 9, 10], Wi-Fi[7], RFID tags[11], acoustic background spectrum[12], geomagnetism[13] or powerlines[14].

Wi-Fi localization. Wi-Fi fingerprinting is performed by collecting the Received Signal Strength

(RSS) for all the available access points (AP) at one location. The RSS can then be normalized or post-processed in any other way before storing it, and the location at which it was recorded in a fingerprint database. A user or device can then use an algorithm to query the fingerprint database and find the approximate location of a fingerprint. The matching can be done either with a probabilistic model by assigning distributions to each data point, or deterministically by using the nearest neighbor algorithm.

Strengths. The main strength of the Wi-Fi fingerprinting approach is the pervasiveness of Wi-Fi. In addtion, most devices nowadays possess Wi-Fi capabilities. This enables the Wi-Fi fingerprinting method to be used in a wide array of devices and locations. **Weaknessess.** One of the common weaknesses is the need to perform a training where data has to be collected from the location in order to build the fingerprint database. This process can be time consuming and tedious. Another weakness is that different Wi-Fi devices measure the RSS differently, and thus they must be calibrated in order to correctly estimate the locations. Also, Wi-Fi signals are not usually very strong, and many APs are required for larger areas.

Outdoor localization

Describe the outdoor localisation technologies, and compare their strengths and weaknesses.

Satellite localization. It works by calculating the pseudo-range based on the speed of light and the propagation delay. Satellites in orbit boradcast the satellite almanac and their clock biases. At least 4 satellites are needed to accurately estimate the position. **Weaknessess.** It provides accurate positioning outdoors with an accuracy of 15-20m. Doesn't work indoors or in urban canyons, and it consumes a lot of energy. **Strengths.** It works all over the world outdoors, one only has to worry about the receiver because the satellites are already in place.

Cellular localization. This system makes use of cellular RF signals to calculate the position. It can estimate the position using several approaches: fingerprinting, angle of arrival, and time of arrival. They can also be used indoors. **Strengths.** Mobile positioning can be used outside or indoors. In most places cellular base stations are present and their exact locations are usually recorded in databases. Cells are far very stable and don't degrade as much over time[10]. Veljo Otsason et al. showed that by using all the detectable mobile cells instead of only the strongest and usable ones the localization system can achieve an accuracy ranging from 2.48m to 5.44m[10]. **Weaknesses.** Accuracy can be low, and can be affected by environmental factors. Cellular base stations are not available in remote areas.

FM localization. Analogously as with cellular signals, FM radio signals can also be used to estimate the position via fingerprinting. Although they can also be used for indoor localization, they also have applications beyond indoors. **Strengths.** Localization using FM fingerprinting benefits from the fact that FM signals travel great distances. FM stations are stable and degrade less over time, thus requiring less calibration[5]. FM signals are also considered to be more energy efficient and are often allowed in sensitive environments[15]. **Weaknessess.** The accuracy can be off by hundreds of meters. Hardware access, or availability, in modern devices to FM signal data is not widespread.

Question 6

Design a conceptual system that includes the technologies you have learnt, for example, wearables, peer-to-peer networks, mobile sensing, Internet of Things, Internet of Nano Things. You may use diagram, UML chart, or any supplementary texts to elaborate.

Bio-factory health system.

In the future, factories may also depend on heavily modified bacteria to produce organic compounds more efficiently. These bio-engineered bacteria would need to be protected, and kept isolated so they don't become contaminated, and their efficiency or byproducts get spoiled. Also, the factories will need to perform actions on a bigger scale, and the Nano things will need to interact with specific machines. This can be done by relying on the Internet of Things.

Components

- Nano bacteria medium/health sensors: in charge of monitoring the bacteria cultivation medium and the quality of the output.
- Nano bacteria 'controllers': in charge of communicating with bacteria and signaling them to change the production rate or trigger self destruction.
- Nano sensor sinks: Mobile sensor network sinks for the data gathered by the nano sensors. They gather it and forward it to the central control.
- Macro machines: These control the lighting, moving, collection of products, and feeding of the bacteria. They are actually a collection of robots that perform most of the 'grunt work'.
- Macro machines mantainance sensors: monitor the state of the machines and alert when maintenance is needed.
- Central control system: Provides oversight over the whole operation and corrects major deviations.

Technologies

- IoT: Internet of Things for the communication with macro machines.
- IoNT: Internet of Nano Things for the communication with nano sensors and nano machines.

- Mobile sensing: this would be used so the nano sensors can report the state of the medium. Macro machine's sensors will be able to communicate the maintenance status of the machines.

By leveraging the Internet of Nano things, we could produce nanosensors that mingle with the bacteria and are able to report the conditions of the bacterias' medium such as nutrient concentrations or foreign bacteria/virus infections. The nano machines can also be used to signal the bacteria to produce or stop producing via molecular communication, or to control the growth rate and production rate of the bacteria.

All the nano sensors would then interact with different machines. Machines that would move the bacteria trays around, light up the bacteria, provide nutrients, process the products of the bacteria, or take action if a bacteria batch becomes contaminated as soon as possible to minimize the impact. This interaction could be provided through wireless sensor networks. Where the nano sensors communicate between each other, and then there is a special kind of nano node that communicates the data via or RF or direct electrical connection.

The machines would also be equipped with sensors to monitor their condition and request maintenance if when required.

Most of the system would be self regulating, but it would be connected and supervised by a central server. The idea is that the communication between the sensors and the different machines is enough to keep the production running. When bigger deviations are detected, the central system would intervene to regulate the environment and production.

The use of nano sensors and nano machines would enable a far more granular control of the bacteria, and it could be leveraged to increase production. One specific bio-factory could be a bio-diesel one.

References

[1] Jukka K. Nurminen. P2P Networks-General. University Lecture, 2014.

[2] Alec Wolman, M Voelker, Nitin Sharma, Neal Cardwell, Anna Karlin, and Henry M Levy. On the scale and performance of cooperative web proxy caching. In *ACM SIGOPS Operating Systems Review*, volume 33, pages 16–31. ACM, 1999.

[3] Doug Beaver, Sanjeev Kumar, Harry C Li, Jason Sobel, Peter Vajgel, et al. Finding a needle in haystack: Facebook's photo storage. In *OSDI*, volume 10, pages 1–8, 2010.

[4] S. Balasubramaniam and J. Kangasharju. Realizing the internet of nano things: Challenges, solutions, and applications. *Computer*, 46(2):62–68, Feb 2013.

[5] Yin Chen, Dimitrios Lymberopoulos, Jie Liu, and Bodhi Priyantha. Fm-based indoor localization. In *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*, MobiSys '12, pages 169–182, New York, NY, USA, 2012. ACM.

[6] Lauri Aalto, Nicklas Göthlin, Jani Korhonen, and Timo Ojala. Bluetooth and wap push based location-aware mobile advertising system. In *Proceedings of the 2Nd International Conference on Mobile Systems, Applications, and Services*, MobiSys '04, pages 49–58, New York, NY, USA, 2004. ACM.

[7] Anthony LaMarca, Yatin Chawathe, Sunny Consolvo, Jeffrey Hightower, Ian Smith, James Scott, Timothy Sohn, James Howard, Jeff Hughes, Fred Potter, Jason Tabert, Pauline Powledge, Gaetano Borriello, and Bill Schilit. Place lab: Device positioning using radio beacons in the wild. In *Proceedings of the Third International Conference on Pervasive Computing*, PERVASIVE'05, pages 116–133, Berlin, Heidelberg, 2005. Springer-Verlag.

[8] H. Laitinen, J. Lahteenmaki, and T. Nordstrom. Database correlation method for GSM location. In *Vehicular Technology Conference, 2001. VTC 2001 Spring. IEEE VTS 53rd*, volume 4, pages 2504–2508 vol.4, 2001.

[9] A. Varshavsky, Anthony LaMarca, Jeffrey Hightower, and E. de Lara. The skyloc floor localization system. In *Pervasive Computing and Communications, 2007. PerCom '07. Fifth Annual IEEE International Conference on*, pages 125–134, March 2007.

[10] Veljo Otsason, Alex Varshavsky, Anthony LaMarca, and Eyal de Lara. Accurate GSM Indoor Localization. In *Proceedings of the 7th International Conference on Ubiquitous Computing*, UbiComp'05, pages 141–158, Berlin, Heidelberg, 2005. Springer-Verlag.

[11] Guang-yao Jin, Xiao-yi Lu, and Myong-Soon Park. An indoor localization mechanism using active rfid tag. In *Proceedings of the IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing -Vol 1 (SUTC'06) - Volume 01*, SUTC '06, pages 40–43, Washington, DC, USA, 2006. IEEE Computer Society.

[12] Stephen P. Tarzia, Peter A. Dinda, Robert P. Dick, and Gokhan Memik. Indoor localization without infrastructure using the acoustic background spectrum. In *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services*, MobiSys '11, pages 155–168, New York, NY, USA, 2011. ACM.

[13] Jaewoo Chung, Matt Donahoe, Chris Schmandt, Ig-Jae Kim, Pedram Razavai, and Micaela Wiseman. Indoor location sensing using geo-magnetism. In *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services*, MobiSys '11, pages 141–154, New York, NY, USA, 2011. ACM.

[14] Erich P. Stuntebeck, Shwetak N. Patel, Thomas Robertson, Matthew S. Reynolds, and Gregory D. Abowd. Wideband powerline positioning for indoor localization. In *Proceedings of the 10th Inter-*

national Conference on Ubiquitous Computing, UbiComp '08, pages 94–103, New York, NY, USA, 2008. ACM.

[15] Aleksandar Matic, Andrei Papliatseyeu, Venet Osmani, and Oscar Mayora-Ibarra. Tuning to your position: FM radio based indoor localization with spontaneous recalibration. In *Pervasive Computing and Communications (PerCom), 2010 IEEE International Conference on*, pages 153–161. IEEE, 2010.