Seminar No.2

Answer the following questions:

- What is the paper about?

This paper describes principles and best practices that Google implements in order to build successful monitoring and alerting systems, and how they are using them for their services.

- What is monitoring?

Monitoring means actually monitoring and controlling the data about a system. It involves all the steps such as collecting, processing, and displaying those data.

- Why monitor a system in the first place?
  Monitoring is essential to identify system failures before they lead to actual problems.
- Explain the 4 golden signals of monitoring.

The 4 golden signals of monitoring are :

1) Latency - In my opinion, it is a very important principle to respect when discussing monitoring because the time to service a request is very important, and distinguishing between a successful request and failed one is crucial
2) Traffic - Traffic also represents an important characteristic to be aware of when monitoring a system, because it measures how much demand is being placed on your system and the decision that you will take.
3) Errors - Monitoring errors is another significant part to be monitored because protocol response codes are insufficient to express all failure conditions.
4) Saturation - It represents how full the service is. It is important to consider these signals also, because as the number of resources is used, the performance degrades, even if there is not 100% of resource utilization.
   If these 4 golden signals tend to be implemented, the service will be at least decently covering by monitoring.

- According to the paper, how do you do the monitoring? What is important? Exemplify.
  In order to perform good monitoring, we should address 2 main questions? What's broken and why. This describes the symptom-causes relationship."What" versus "why" is one of the most important distinctions in writing good monitoring with maximum signal and minimum noise.
- What approach would you use for your lab: White-box or Black-box monitoring? Why?

In my opinion, for our lab I will go with White-box monitoring, I don't know how my colleague will choose, and let me explain why. This type of monitoring mainly refers to monitoring the internal states of the applications running on our system. Mainly this type of monitoring involves exposing metrics that are specific to your application like the total number of HTTP requests received / latency etc. I think that we will go better with this approach because it allows detection of imminent problems, failures masked by retries, and so forth.

- What happened with Bigtable SRE and how did they "fix" the situation?

Many years ago, the Bigtable service's SLO was based on a synthetic well-behaved client's mean performance. Because of problems in Bigtable and lower layers of the storage stack, the mean performance was driven by a "large" tail: the worst 5% of requests were often significantly slower than the rest.

To remedy the situation, the team used a three-pronged approach: while making great efforts to improve the performance of Bigtable, we also temporarily dialed back our SLO target, using the 75th percentile request latency. We also disabled email alerts, as there were so many that spending time diagnosing them was infeasible. This strategy gave us enough breathing room to actually fix the longer-term problems in Bigtable and the lower layers of the storage stack, rather than constantly fixing tactical problems.