Michael Salib 6.033 2/22/02

The AMPED architecture will outperform the SPED architecture because the relative scarcity of RAM compared to disk space in many systems will remain despite shrinking RAM costs. That scarcity results from both downward trends in the cost of disk space as well as relative increases in memory demand. Even without such scarcity, several practical aspects of web serving will reduce any apparent performance advantage that SPED has.

The relative scarcity of RAM compared to disk space will ensure that AMPED provides better performance. If the past is any guide, RAM costs will decrease exponentially, but so will disk costs. Past experience suggests that the ratio between the cost of one MB of disk versus one MB of RAM will not change, even as the individual prices dramatically shrink. This means that for a limited budget, system designers will still have to choose between RAM and disk. Given the choice between durable disk storage and volatile ram, designers will prefer disk to RAM, producing systems where the amount of RAM relative to disk is seriously constrained even though the absolute cost of RAM is low.

The growth in demand for memory will also reduce the amount of available RAM relative to disk space, further penalizing the SPED architecture. Just because one has a 100 terabytes of RAM doesn't mean they have nothing better to spend it on than caching web data. Regardless of the cost of memory, there will never be enough of it compared to what system designers want. As the cost of memory decreases, the demands for ever larger memories will only increase. In fact, as memory cost decrease, whole new types of applications become possible. These new applications all share an insatiable hunger for memory, a hunger that can only be satisfied by starving less important applications, like http servers.

Even in environments where RAM is plentiful compared to disk, the dynamic nature of web serving ensures that AMPED will perform very nearly as well as SPED architectures. The authors claim that the performance difference between the two architectures results from overhead imposed by the AMPED server checking for the existence of data in the cache. That deficiency only applies when serving static data; since dynamic data can't be cached, the performance difference between the two architectures when serving dynamic data should be negligible. In fact, the majority of data that web servers transmit is dynamically generated. This effect reduces the already small distinction between architecture performance even further. In addition, many sites place dedicated proxy engines between their web servers and clients. This practice dramatically reduces the need for caching by the web server itself thus further reducing any performance advantage the SPED architecture had over AMPED. The authors of the Flash paper neglected to consider either effect.