Lecture 4#

System Performance Tuning

System Performance Tuning

- System performance tuning is a complex subject, in which no part of the system is left.
- Although it is quite easy to pin-point general performance problems, it is harder to make general recommendations to fix these.

1. What processes are running
2. How much available memory the system has
3. Whether disks are being used excessively
4. Whether the network is being used heavily
5. What software dependencies the system has (e.g. DNS, NFS).

1.1Resources and dependencies

- Since all resources are scheduled by processes, it is natural to check the process table first and then look at resource usage
- On Windows, one has the process manager and performance monitor for this. On Unix-like systems, we check the process listing with ps aux

```
host% ps aux | more

USER    PID %CPU %MEM   SZ  RSS TT       S    START  TIME COMMAND
root      3  0.2  0.0    0    0 ?        S   Jun 15 55:38 fsflush
root  22112  0.1  0.5 1464 1112 pts/2    O 15:39:54  0:00 ps aux
mark  22113  0.1  0.3 1144  720 pts/2    O 15:39:54  0:00 more
root    340  0.1  0.4 1792  968 ?        S   Jun 15  3:13 /bin/fingerd
```

- This one was taken on a quiet system, with no load.
- The columns show the user ID of the process, the process ID, an indication of the amount of CPU time used in executing the program and an indication of the amount of memory allocated.
- The SZ post is the size of the process in total (code plus data plus stack), while RSS is the resident size, or how much of the program code is actually resident in RAM, as opposed to being paged out, or never even loaded. TIME shows the amount of CPU time accumulated by the process, while START indicates the amount of clock time which has elapsed since the process started

- Problem processes are usually identified by:
- • %CPU is large. A CPU-intensive process or a process which has gone into an endless loop. TIME is large. A program which has been CPU intensive, or which has been stuck in a loop for a long period.
- • %MEM is large. SZ is large. A large and steadily growing value can indicate a memory leak.

- Unix-like systems also tell us about memory performance through the virtual memory statistics, e.g. the **vmstat command**. This command gives a different output on each operating system, but summarizes the amount of free memory as well as paging performance etc

OS	List virtual memory usage
AIX	`lsps -a`
HPUX	`swapinfo -t -a -m`
Digital Unix/OSF1	`swapon -s`
Solaris 1 or SunOS 3/4	`pstat -s`
Solaris 2 or SunOS 5	`swap -l`
GNU/Linux	`free`
Windows	Performance manager

- **Excessive network traffic** is also a cause of **impaired performance**. We should try to eliminate unnecessary network traffic whenever possible

1. Make sure that there is a DNS server on each large subnet to avoid sending unnecessary queries through a router.
2. Make sure that the nameservers themselves use the loopback address 127.0.0.1 as the primary nameserver on Unix-like hosts, so that we do not cause collisions by having the nameserver talk to itself on the public network.
3. Try to avoid distributed file accesses on a different subnet. This loads the router. If possible, file-servers and clients should be on the same subnet.

- netstat, - graphical tools for viewing network statistics
- Once a problem is identified, we need a strategy for solving it. Performance tuning can involve everything from changing hardware to tweaking software.
- • Optimizing choice of hardware
- • Optimizing chosen hardware
- • Optimizing kernel behavior
- • Optimizing software configurations

1.2 Hardware

- **Disks**
- **Network**
- **Ethernet collisions**
- **Disk thrashing**

Disks

- When assigning partitions to new disks, it pays to use the fastest disks for the data which are accessed most often, e.g. for user home directories.
- To improve disk performance, we can do two things.

1. buy faster disks
2. parallelism to overcome the time it takes for physical motions to be executed.

- The mechanical problem which is inherent in disk drives is that the heads which read and write data have to move as a unit. If we need to collect two files concurrently which lie spread all over the disk, this has to be done serially.

- Disk striping is a technique whereby file systems are spread over several disks. By spreading files over several disks, we have several sets of disk heads which can seek independently of one another, and work in parallel.
- This does not necessarily increase the transfer rate, but it does lower seek times, and thus performance improvement can approach as much as N times with N disks.
- RAID technologies employ striping techniques and are widely available commercially. Spreading disks and files across multiple disk controllers will also increase parallelism

Network

- Performance of network is measured by 2 characteristics:-
- **Latency**:- the time delay its takes messages to be transported across the network
- **Throughput**:- the rate at which the message is transmitted across the network

What affects network system performance

(sending message from one system to other across network)

- **1) host system (CPU/Memory)**
- Does hosts system have sufficient processors and memory resources to support the desired applications and operating system network stacks?
- **2) host system (interconnected)**
- Does the interconnection between hosts system and NIC have sufficient BW
- What latency does the interconnect add?

- **3) NIC**
- Does NIC have sufficient resources to transmits or receive all packets requested of its ?
- What latency does NIC add?
- **4) Network:**
- What is the raw channel capacity (BW) of the network wire?
- Does all network devices router/switches have sufficient BW?
- What latency does router/switches add ?
- What is latency of wire?

Ethernet collisions

- The Ethernet cable is a shared bus. When a host wishes to communicate with another host, it simply tries. If another host happens to be using the bus at that time, there is a collision and the host must try again at random until it is heard. This method naturally leads to contention for bandwidth.
- The system works **quite well** when **traffic is low**, but as the **number of hosts** competing for **bandwidth increases**, the probability of a **collision increases in step**.
- **Contention can only be reduced** by **reducing the amount of traffic on the network segment.**

Disk thrashing

- **Thrashing is a problem** which occurs **because** of the **slowness of disk head movements**, **compared** with the **speed of kernel time-sharing algorithms**.
- **If two processes attempt to take control of a resource simultaneously, the kernel and its device drivers attempt to minimize the motion of the heads by queuing requested blocks in a special order.**
- The algorithms really try to make the disks traverse the disk platter uniformly, but the requests do not always come in a predictable or congenial order.
- **The result is that the disk heads can be forced back and forth across the disk, driven by different processes and slowing the system to a virtual standstill.**

- An even worse situation can arise with the virtual memory system. If a host begins paging to disk because it is low on memory, then there can be simultaneous contention both for memory and for disk.
- Imagine, for instance, that there are many processes, each loading files into memory, when there is no free RAM.
- In order to use RAM, some has to be freed by paging to disk; but the disk is already busy seeking files.
- In order to load a file, memory has to be freed, but memory can't be freed until the disk is free to page, this drags the heads to another partition, then back again ... and so on.
- This nightmare brings the system to a virtual standstill as it fights both over free RAM and disk head placement.
- **The only cure for thrashing is to increase memory, or reduce the number of processes contending for resources.**

1.3 Software tuning and kernel configuration

- **Software performance tuning** is a more **complex problem than hardware performance tuning,** simply because the options we have for tuning software depend on what the software is, how it is written and whether or not the designer made it easy for us to tune its performance
- **Performance tuning is related to the availability** or **sharing of system resources**. This requires **tuning the system kernel**. The most **configurable piece of software** on the **system is the kernel.**

- many kernel parameters can be set at run time using the kernel module configuration command **ndd.**
- Others can be configured in a single file /etc/system. The parameters in this file can be set with a reboot of the kernel, using the reconfigure flag
- **reboot -- -r**
- For instance, on a heavily loaded system which allows many users to run external logins, terminals, or X-terminal software, we need to increase many of the default system parameters
- The **file /etc/system**, then looks like this:
- **set maxusers=100**
- **Most Unix-like operating systems do not permit run-time configuration. New kernels have to be compiled and the values hard-coded into the kernel. This requires not just a reboot, but a recompilation of the kernel in order to make a change**

1.4 Data efficiency

- Efficiency of storage and transmission depends on the configuration parameters used to manage disks and networks, and also on the amount of traffic the devices
- Some filesystem formatting programs on Unix-like systems allow us to reserve a certain percentage of disk space for privileged users.
- For instance, the default for BSD is to reserve ten percent of the size of a partition for use by privileged processes only. The idea here is to prevent the operating system from choking due to the activities of users.
- If we have partitioned a host so as to separate users from the operating system, then there is no need to reserve space on user disks.

- Another issue with disk efficiency is the configuration of block sizes
- Blocks are quite large, usually around 8 kilobytes. Even if we allocate a file which is one byte long, it will be stored as a separate unit, in a block by itself, or in a fragment. Fragments are usually around 1 kilobyte
- If we have many small files, this can clearly lead to a large wastage of space and it might lead to decrease the filesystem block size
- If, we deal with mostly large files, then the block size could be increased to improve transfer efficiency
- The filesystem parameters can, in other words, be tuned to balance file size and transfer-rate efficiency. Normally the default settings are a good compromise