Performance tuning the Linux Kernel on the ARM Architecture

(Playing with a cell phone!)

Primary Objective: Preventative Maintenance

- Attempting to foresee probable bottle-necks and detect areas of poor performance during the very early development stages
- We have the advantage that we are very early in the development of the cluster, which gives us a window of opportunity to change tuning parameters without worrying about denial of service conditions

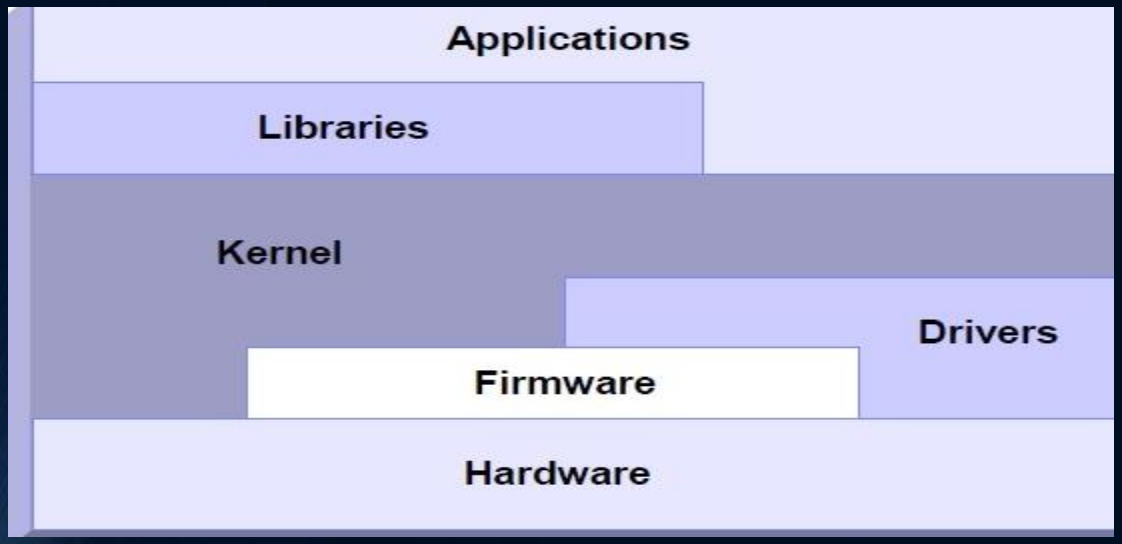
"Optimising the linux environment for higher performance usually takes place once a working system is running to specification and unforeseen bottlenecks or bugs occur"

-Linux Performance and Tuning Guidelines, (RedPaper)

ibm.com/redbooks

System
Kernel
Linux 3.4.61-sun7i+ arm7l

Tuning Parameters



AllWinnerTech SOC A20, ARM® Cortex™-A7 Dual-Core ARM® Mali400 MP2 Complies with OpenGL ES 2.0/1.1
1GB DDR3 @480M

What do we tune for? Application specific

- Higher performance configuration will skew system for the intended workload
- Expect poor performance for different workload patterns
- We are not sure at present - IO intensive?
- We want to understand what parameters are relevant for our purposes

net.ipv4.conf.default.arp_accept = 0
net.ipv4.conf.default.arp_announce = 0
net.ipv4.conf.default.arp_filter = 0
net.ipv4.conf.default.arp_ignore = 0
net.ipv4.conf.default.arp_notify = 0
net.ipv4.conf.default.bootp_relay = 0
net.ipv4.conf.default.disable_policy = 0
net.ipv4.conf.default.disable_xfrm = 0
net.ipv4.conf.default.force_igmp_version = 0
net.ipv4.conf.default.forwarding = 0
net.ipv4.conf.default.log_martians = 0
net.ipv4.conf.default.mc_forwarding = 0
net.ipv4.conf.default.medium_id = 0
net.ipv4.conf.default.promote_secondaries = 0
net.ipv4.conf.default.proxy_arp = 0
net.ipv4.conf.default.proxy_arp_pvlan = 0
net.ipv4.conf.default.rp_filter = 1
net.ipv4.conf.default.secure_redirects = 1
net.ipv4.conf.default.send_redirects = 1
net.ipv4.conf.default.shared_media = 1
net.ipv4.conf.default.src_valid_mark = 0
net.ipv4.conf.default.tag = 0
net.ipv4.conf.eth0.accept_local = 0
net.ipv4.conf.eth0.accept_redirects = 1
net.ipv4.conf.eth0.accept_source_route = 1
net.ipv4.conf.eth0.arp_accept = 0
net.ipv4.conf.eth0.arp_announce = 0
net.ipv4.conf.eth0.arp_filter = 0
net.ipv4.conf.eth0.arp_ignore = 0
net.ipv4.conf.eth0.arp_notify = 0
net.ipv4.conf.eth0.bootp_relay = 0
net.ipv4.conf.eth0.disable_policy = 0
net.ipv4.conf.eth0.disable_xfrm = 0
net.ipv4.conf.eth0.force_igmp_version = 0
net.ipv4.conf.eth0.forwarding = 0
net.ipv4.conf.eth0.log_martians = 0
net.ipv4.conf.eth0.mc_forwarding = 0
net.ipv4.conf.eth0.medium_id = 0
net.ipv4.conf.eth0.promote_secondaries = 0
net.ipv4.conf.eth0.proxy_arp = 0
net.ipv4.conf.eth0.proxy_arp_pvlan = 0
net.ipv4.conf.eth0.rp_filter = 1
net.ipv4.conf.eth0.secure_redirects = 1
net.ipv4.conf.eth0.send_redirects = 1
net.ipv4.conf.eth0.shared_media = 1
net.ipv4.conf.eth0.src_valid_mark = 0
net.ipv4.conf.eth0.tag = 0
net.ipv4.conf.lo.accept_local = 0
net.ipv4.conf.lo.accept_redirects = 1
net.ipv4.conf.lo.accept_source_route = 1
net.ipv4.conf.lo.arp_accept = 0
net.ipv4.conf.lo.arp_announce = 0
net.ipv4.conf.lo.arp_filter = 0
net.ipv4.conf.lo.arp_ignore = 0
net.ipv4.conf.lo.arp_notify = 0
net.ipv4.conf.lo.bootp_relay = 0
net.ipv4.conf.lo.disable_policy = 1
net.ipv4.conf.lo.disable_xfrm = 1
net.ipv4.conf.lo.force_igmp_version = 0
net.ipv4.conf.lo.forwarding = 0
net.ipv4.conf.lo.log_martians = 0
net.ipv4.conf.lo.mc_forwarding = 0
net.ipv4.conf.lo.medium_id = 0
net.ipv4.conf.lo.promote_secondaries = 0
net.ipv4.conf.lo.proxy_arp = 0
net.ipv4.conf.lo.proxy_arp_pvlan = 0
net.ipv4.conf.lo.rp_filter = 1
net.ipv4.conf.lo.secure_redirects = 1
net.ipv4.conf.lo.send_redirects = 1
net.ipv4.conf.lo.shared_media = 1
net.ipv4.conf.lo.src_valid_mark = 0
net.ipv4.conf.lo.tag = 0
net.ipv4.conf.tunl0.accept_local = 0
net.ipv4.conf.tunl0.accept_redirects = 1
net.ipv4.conf.tunl0.accept_source_route = 1
net.ipv4.conf.tunl0.arp_accept = 0
net.ipv4.conf.tunl0.arp_announce = 0
net.ipv4.conf.tunl0.arp_filter = 0
net.ipv4.conf.tunl0.arp_ignore = 0
net.ipv4.conf.tunl0.arp_notify = 0
net.ipv4.conf.tunl0.bootp_relay = 0
net.ipv4.conf.tunl0.disable_policy = 0
net.ipv4.conf.tunl0.disable_xfrm = 0
net.ipv4.conf.tunl0.force_igmp_version = 0
net.ipv4.conf.tunl0.forwarding = 0
net.ipv4.conf.tunl0.log_martians = 0
net.ipv4.conf.tunl0.mc_forwarding = 0
net.ipv4.conf.tunl0.medium_id = 0
net.ipv4.conf.tunl0.promote_secondaries = 0
net.ipv4.conf.tunl0.proxy_arp = 0
net.ipv4.conf.tunl0.proxy_arp_pvlan = 0
net.ipv4.conf.tunl0.rp_filter = 1
net.ipv4.conf.tunl0.secure_redirects = 1
net.ipv4.conf.tunl0.send_redirects = 1
net.ipv4.conf.tunl0.shared_media = 1
net.ipv4.conf.tunl0.src_valid_mark = 0
net.ipv4.conf.tunl0.tag = 0
net.ipv4.icmp_echo_ignore_all = 0
net.ipv4.icmp_echo_ignore_broadcasts = 1
net.ipv4.icmp_errors_use_inbound_ifaddr = 0
net.ipv4.icmp_ignore_bogus_error_responses = 1
net.ipv4.icmp_ratelimit = 1000
net.ipv4.icmp_ratemask = 6168
net.ipv4.igmp_max_memberships = 20
net.ipv4.igmp_max_msf = 10
net.ipv4.inet_peer_maxttl = 600
net.ipv4.inet_peer_minttl = 120
net.ipv4.inet_peer_threshold = 65664
net.ipv4.ip_default_ttl = 64
net.ipv4.ip_dynaddr = 0
net.ipv4.ip_forward = 0
net.ipv4.ip_local_port_range = 32768 61000
net.ipv4.ip_local_reserved_ports =
net.ipv4.ip_no_pmtu_disc = 0
net.ipv4.ip_nonlocal_bind = 0
net.ipv4.ipfrag_high_thresh = 262144
net.ipv4.ipfrag_low_thresh = 196608
net.ipv4.ipfrag_max_dist = 64
net.ipv4.ipfrag_secret_interval = 600
net.ipv4.ipfrag_time = 30
net.ipv4.neigh.default.anycast_delay = 100
net.ipv4.neigh.default.app_solicit = 0
net.ipv4.neigh.default.base_reachable_time_ms = 30000
net.ipv4.neigh.default.delay_first_probe_time = 5
net.ipv4.neigh.default.gc_interval = 30
net.ipv4.neigh.default.gc_stale_time = 60
net.ipv4.neigh.default.gc_thresh1 = 128
net.ipv4.neigh.default.gc_thresh2 = 512
net.ipv4.neigh.default.gc_thresh3 = 1024
net.ipv4.neigh.default.locktime = 100
net.ipv4.neigh.default.mcast_solicit = 3
net.ipv4.neigh.default.proxy_delay = 80
net.ipv4.neigh.default.proxy_qlen = 64
net.ipv4.neigh.default.retrans_time_ms = 1000
net.ipv4.neigh.default.ucast_solicit = 3
net.ipv4.neigh.default.unres_qlen = 35
net.ipv4.neigh.default.unres_qlen_bytes = 65536
net.ipv4.neigh.eth0.anycast_delay = 100
net.ipv4.neigh.eth0.app_solicit = 0
net.ipv4.neigh.eth0.base_reachable_time_ms = 30000
net.ipv4.neigh.eth0.delay_first_probe_time = 5
net.ipv4.neigh.eth0.gc_stale_time = 60
net.ipv4.neigh.eth0.locktime = 100
net.ipv4.neigh.eth0.mcast_solicit = 3
net.ipv4.neigh.eth0.proxy_delay = 80
net.ipv4.neigh.eth0.proxy_qlen = 64
net.ipv4.neigh.eth0.retrans_time_ms = 1000
net.ipv4.neigh.eth0.ucast_solicit = 3
net.ipv4.neigh.eth0.unres_qlen = 35
net.ipv4.neigh.eth0.unres_qlen_bytes = 65536
net.ipv4.neigh.lo.gc_stale_time = 60
net.ipv4.neigh.lo.locktime = 100
net.ipv4.neigh.lo.mcast_solicit = 3
net.ipv4.neigh.lo.proxy_delay = 80
net.ipv4.neigh.lo.proxy_qlen = 64
net.ipv4.neigh.lo.retrans_time_ms = 1000
net.ipv4.neigh.lo.ucast_solicit = 3
net.ipv4.neigh.lo.unres_qlen = 35
net.ipv4.neigh.lo.unres_qlen_bytes = 65536
net.ipv4.neigh.tunl0.anycast_delay = 100
net.ipv4.neigh.tunl0.app_solicit = 0
net.ipv4.neigh.tunl0.base_reachable_time_ms = 30000
net.ipv4.neigh.tunl0.delay_first_probe_time = 5
net.ipv4.neigh.tunl0.gc_stale_time = 60
net.ipv4.neigh.tunl0.locktime = 100
net.ipv4.neigh.tunl0.mcast_solicit = 3
net.ipv4.neigh.tunl0.proxy_delay = 80
net.ipv4.neigh.tunl0.proxy_qlen = 64
net.ipv4.neigh.tunl0.retrans_time_ms = 1000
net.ipv4.neigh.tunl0.ucast_solicit = 3
net.ipv4.neigh.tunl0.unres_qlen = 35
net.ipv4.neigh.tunl0.unres_qlen_bytes = 65536
net.ipv4.ping_group_range = 1 0
net.ipv4.route.error_burst = 500
net.ipv4.route.error_cost = 100
net.ipv4.route.gc_elasticity = 8
net.ipv4.route.gc_interval = 60
net.ipv4.route.gc_min_interval = 0
net.ipv4.route.gc_min_interval_ms = 500
net.ipv4.route.gc_thresh = 32768
net.ipv4.route.gc_timeout = 300
net.ipv4.route.max_size = 524288
net.ipv4.route.min_adv_mss = 256
net.ipv4.route.min_pmtu = 552
net.ipv4.route.mtu_expires = 600
net.ipv4.route.redirect_load = 2
net.ipv4.route.redirect_number = 9
net.ipv4.route.redirect_silence = 2048
net.ipv4.rt_cache_rebuild_count = 4
net.ipv4.tcp_abc = 0
net.ipv4.tcp_abort_on_overflow = 0
net.ipv4.tcp_adv_win_scale = 1
net.ipv4.tcp_allowed_congestion_control = cubic reno
net.ipv4.tcp_app_win = 31
net.ipv4.tcp_available_congestion_control = cubic reno bic westwood highspeed hybla htcp vegas veno scalable lp yeah illinois
net.ipv4.tcp_base_mss = 512
net.ipv4.tcp_challenge_ack_limit = 100
net.ipv4.tcp_congestion_control = cubic
net.ipv4.tcp_cookie_size = 0
net.ipv4.tcp_dsack = 1
net.ipv4.tcp_ecn = 2
net.ipv4.tcp_fack = 1
net.ipv4.tcp_fin_timeout = 60
net.ipv4.tcp_frto = 2
net.ipv4.tcp_frto_response = 0
net.ipv4.tcp_keepalive_intvl = 75
net.ipv4.tcp_keepalive_probes = 9
net.ipv4.tcp_keepalive_time = 7200
net.ipv4.tcp_latency = 0
net.ipv4.tcp_max_orphans = 65536
net.ipv4.tcp_max_ssthresh = 0
net.ipv4.tcp_max_syn_backlog = 512
net.ipv4.tcp_max_tw_buckets = 65536
net.ipv4.tcp_mem = 17973 23965 35946
net.ipv4.tcp_moderate_rcvbuf = 1
net.ipv4.tcp_mtu_probing = 0
net.ipv4.tcp_no_metrics_save = 0
net.ipv4.tcp_orphan_retries = 0
net.ipv4.tcp_reordering = 3
net.ipv4.tcp_retrans_collapse = 1
net.ipv4.tcp_retries1 = 3
net.ipv4.tcp_retries2 = 15
net.ipv4.tcp_rfc1337 = 0
net.ipv4.tcp_rmem = 4096 87380 6177280
net.ipv4.tcp_sack = 1
net.ipv4.tcp_slow_start_after_idle = 1
net.ipv4.tcp_stdurg = 0
net.ipv4.tcp_syn_retries = 5
net.ipv4.tcp_synack_retries = 5
net.ipv4.tcp_syncookies = 1
net.ipv4.tcp_thin_dupack = 0
net.ipv4.tcp_thin_linear_timeouts = 0
net.ipv4.tcp_timestamps = 1
net.ipv4.tcp_tso_win_divisor = 3
net.ipv4.tcp_tw_recycle = 0
net.ipv4.tcp_tw_reuse = 0
net.ipv4.tcp_window_scaling = 1
net.ipv4.tcp_wmem = 4096 16384 4194304
net.ipv4.tcp_workaround_signed_windows = 0
net.ipv4.udp_mem = 18096 24130 36192
net.ipv4.udp_rmem_min = 4096
net.ipv4.udp_wmem_min = 4096
net.ipv4.xfrm4_gc_thresh = 262144
net.ipv6.bindv6only = 0
net.ipv6.conf.all.accept_dad = 1
net.ipv6.conf.all.accept_ra = 1
net.ipv6.conf.all.accept_ra_defrtr = 1
net.ipv6.conf.all.accept_ra_pinfo = 1
net.ipv6.conf.all.accept_ra_rt_info_max_plen = 0
net.ipv6.conf.all.accept_ra_rtr_pref = 1
net.ipv6.conf.all.accept_redirects = 1
net.ipv6.conf.all.accept_source_route = 0
net.ipv6.conf.all.autoconf = 1
net.ipv6.conf.all.dad_transmits = 1
net.ipv6.conf.all.disable_ipv6 = 0
net.ipv6.conf.all.force_mld_version = 0
net.ipv6.conf.all.force_tllao = 0
net.ipv6.conf.all.forwarding = 0
net.ipv6.conf.all.hop_limit = 64
net.ipv6.conf.all.max_addresses = 16
net.ipv6.conf.all.max_desync_factor = 600
net.ipv6.conf.all.mc_forwarding = 0
net.ipv6.conf.all.mtu = 1280
net.ipv6.conf.all.optimistic_dad = 0
net.ipv6.conf.all.proxy_ndp = 0
net.ipv6.conf.all.regen_max_retry = 3
net.ipv6.conf.all.router_probe_interval = 60
net.ipv6.conf.all.router_solicitation_delay = 1
net.ipv6.conf.all.router_solicitation_interval = 4
net.ipv6.conf.all.router_solicitations = 3
net.ipv6.conf.all.temp_prefered_lft = 86400
net.ipv6.conf.all.temp_valid_lft = 604800
net.ipv6.conf.all.use_tempaddr = 2
net.ipv6.conf.default.accept_dad = 1
net.ipv6.conf.default.accept_ra = 1
net.ipv6.conf.default.accept_ra_defrtr = 1
net.ipv6.conf.default.accept_ra_pinfo = 1
net.ipv6.conf.default.accept_ra_rt_info_max_plen = 0
net.ipv6.conf.default.accept_ra_rtr_pref = 1
net.ipv6.conf.default.accept_redirects = 1
net.ipv6.conf.default.accept_source_route = 0
net.ipv6.conf.default.autoconf = 1
net.ipv6.conf.default.dad_transmits = 1
net.ipv6.conf.default.disable_ipv6 = 0
net.ipv6.conf.default.force_mld_version = 0
net.ipv6.conf.default.force_tllao = 0
net.ipv6.conf.default.forwarding = 0
net.ipv6.conf.default.hop_limit = 64
net.ipv6.conf.default.max_addresses = 16
net.ipv6.conf.default.max_desync_factor = 600
net.ipv6.conf.default.mc_forwarding = 0
net.ipv6.conf.default.mtu = 1280
net.ipv6.conf.default.optimistic_dad = 0
net.ipv6.conf.default.proxy_ndp = 0
net.ipv6.conf.default.regen_max_retry = 3
net.ipv6.conf.default.router_probe_interval = 60
net.ipv6.conf.default.router_solicitation_delay = 1
net.ipv6.conf.default.router_solicitation_interval = 4
net.ipv6.conf.default.router_solicitations = 3
net.ipv6.conf.default.temp_prefered_lft = 86400
net.ipv6.conf.default.temp_valid_lft = 604800
net.ipv6.conf.default.use_tempaddr = 2
net.ipv6.conf.eth0.accept_dad = 1
net.ipv6.conf.eth0.accept_ra = 1
net.ipv6.conf.eth0.accept_ra_defrtr = 1
net.ipv6.conf.eth0.accept_ra_pinfo = 1
net.ipv6.conf.eth0.accept_ra_rt_info_max_plen = 0
net.ipv6.conf.eth0.accept_ra_rtr_pref = 1
net.ipv6.conf.eth0.accept_redirects = 1
net.ipv6.conf.eth0.accept_source_route = 0
net.ipv6.conf.eth0.autoconf = 1
net.ipv6.conf.eth0.dad_transmits = 1
net.ipv6.conf.eth0.disable_ipv6 = 0
net.ipv6.conf.eth0.force_mld_version = 0
net.ipv6.conf.eth0.force_tllao = 0
net.ipv6.conf.eth0.forwarding = 0
net.ipv6.conf.eth0.hop_limit = 64
net.ipv6.conf.eth0.max_addresses = 16
net.ipv6.conf.eth0.max_desync_factor = 600
net.ipv6.conf.eth0.mc_forwarding = 0
net.ipv6.conf.eth0.mtu = 1500
net.ipv6.conf.eth0.optimistic_dad = 0

794

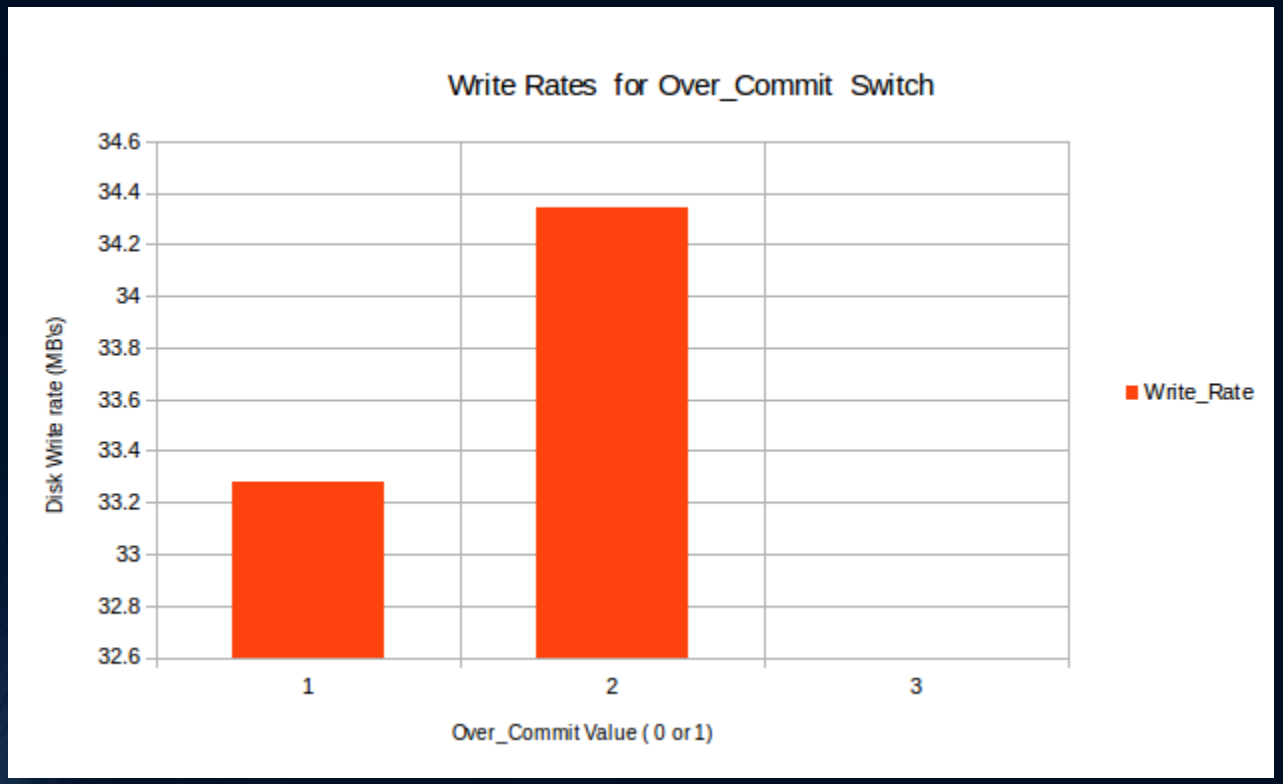
How do we tune?

/proc/sys

/proc/sys/fs/*	Used to increase the number of open files the OS allows and to handle quota.
/proc/sys/kernel/*	For tuning purposes, you can enable hotplug, manipulate shared memory, and specify the maximum number of PID files and level of debug in syslog.
/proc/sys/net/*	Tuning of network in general, IPV4 and IPV6.
/proc/sys/vm/*	Management of cache memory and buffer.

```
/$ sysctl -a
/$ sysctl vm > /home/jonathan/vm.txt
/$ sysctl   vm.overcommit_memory
/$ 0
/$ sysctl  -w vm.overcommit_memory=1
/$1
```

Diving In: dd command

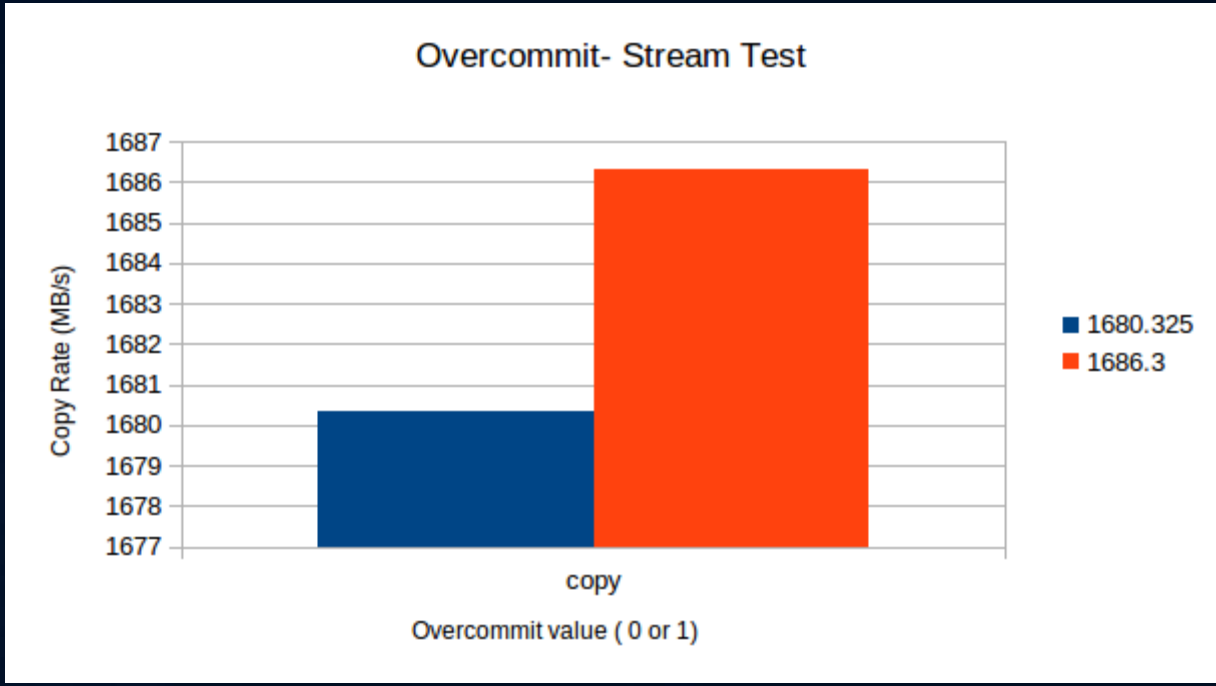


3% improvement in disk write rate

.overcommit_memory=1

The kernel will grant every memory allocation without checking whether the allocation can be satisfied

Stream results for overcommit=1



Stream – popular benchmark for memory

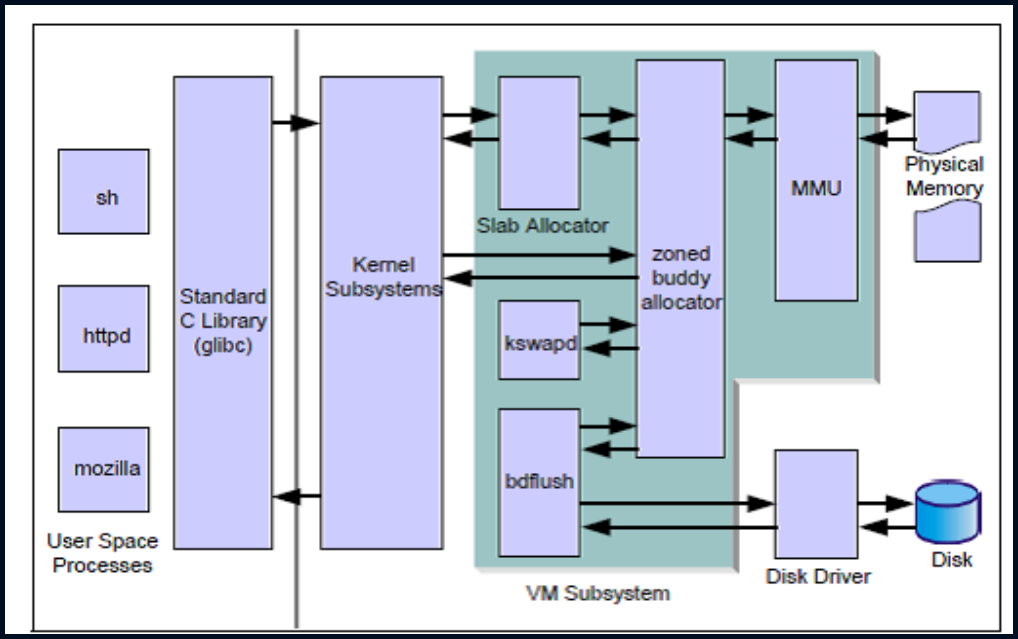
Four functions

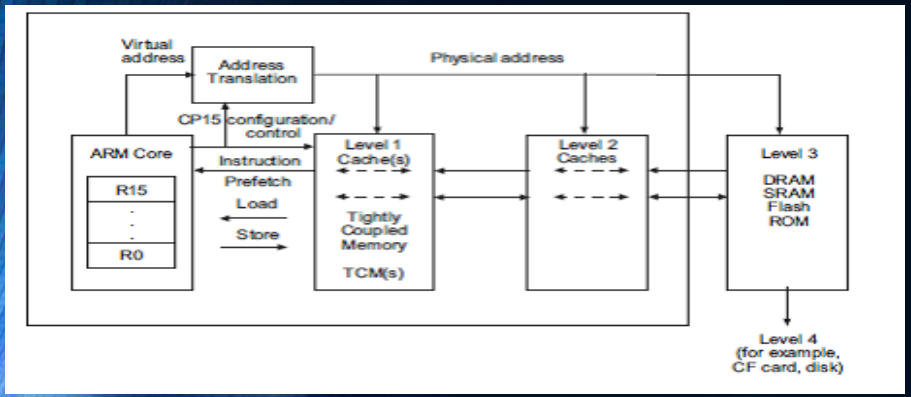
1. Copy: A(i) = b(i)
2. Add: a(i) = b(i) + c(i)
3. Scale: a(i) = q(bi)
4. Triad: a(i) = b(i) + q(ci)

/proc/sys/vm

vm.block_dump = 0
vm.dirty_background_bytes = 0
vm.dirty_background_ratio = 10
vm.dirty_bytes = 0
vm.dirty_expire_centisecs = 3000
vm.dirty_ratio = 20
vm.dirty_writeback_centisecs = 500
vm.drop_caches = 0
vm.extfrag_threshold = 500
vm.extra_free_kbytes = 0
vm.highmem_is_dirtyable = 0
vm.laptop_mode = 0
vm.legacy_va_layout = 0
vm.lowmem_reserve_ratio = 256 32 32
vm.max_map_count = 65530
vm.min_free_kbytes = 16384
vm.min_free_order_shift = 1
vm.mmap_min_addr = 32768
vm.nr_pdflush_threads = 0
vm.oom_dump_tasks = 1
vm.oom_kill_allocating_task = 0
vm.overcommit_memory = 1
vm.overcommit_ratio = 50
vm.page-cluster = 3
vm.panic_on_oom = 0
vm.percpu_pagelist_fraction = 0
vm.scan_unevictable_pages = 0
vm.stat_interval = 1
vm.swappiness = 60
vm.vfs_cache_pressure = 100

Where to from here?





1	vm.block_dump	=	0
2	vm.dirty_background_bytes	=	0
3	vm.dirty_background_ratio	=	5
4	vm.dirty_bytes	=	0
5	vm.dirty_expire_centisecs	=	3000
6	vm.dirty_ratio	=	10
7	vm.dirty_writeback_centisecs	=	500
8	vm.drop_caches	=	0
9	vm.extfrag_threshold	=	500
10	vm.hugepages_treat_as_movable	=	0
11	vm.hugetlb_shm_group	=	0
12	vm.laptop_mode	=	0
13	vm.legacy_va_layout	=	0
14	vm.lowmem_reserve_ratio	=	256
15	vm.max_map_count	=	65530
16	vm.memory_failure_early_kill	=	0
17	vm.memory_failure_recovery	=	1
18	vm.min_free_kbytes	=	67584
19	vm.min_slab_ratio	=	5

20	vm.min_unmapped_ratio	=	1
21	vm.mmap_min_addr	=	65536
22	vm.nr_hugepages	=	0
24	vm.nr_overcommit_hugepages	=	0
25	vm.nr_pdflush_threads	=	0
26	vm.numa_zonelist_order	=	default
27	vm.oom_dump_tasks	=	1
28	vm.oom_kill_allocating_task	=	0
29	vm.overcommit_memory	=	0
30	vm.overcommit_ratio	=	50
31	vm.page-cluster	=	3
32	vm.panic_on_oom	=	0
33	vm.percpu_pagelist_fraction	=	0
34	vm.scan_unevictable_pages	=	0
35	vm.stat_interval	=	1
36	vm.swappiness	=	60
37	vm.vfs_cache_pressure	=	100
38	vm.zone_reclaim_mode	=	0

Flop/w

Thank You