Software-level Power-Aware Computing

Lecture 3

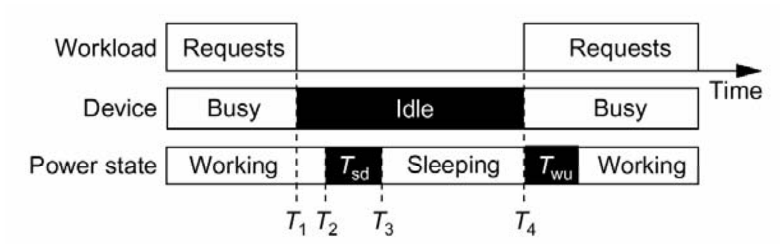
Lecture Organizations

- Lecture 1:
  - Introduction to Low-power systems
  - Low-power binary encoding
  - Power-aware compiler techniques
- Lectures 2 & 3
  - Dynamic voltage scaling (DVS) techniques
    - OS-level DVS: Inter-Task DVS
    - Compiler-level DVS: Intra-Task DVS
    - Application-level DVS
  - Dynamic power management
- Lecture 4
  - Software power estimation & optimization
  - Low-power techniques for multiprocessor systems
  - Leakage reduction techniques

Dynamic Power Management

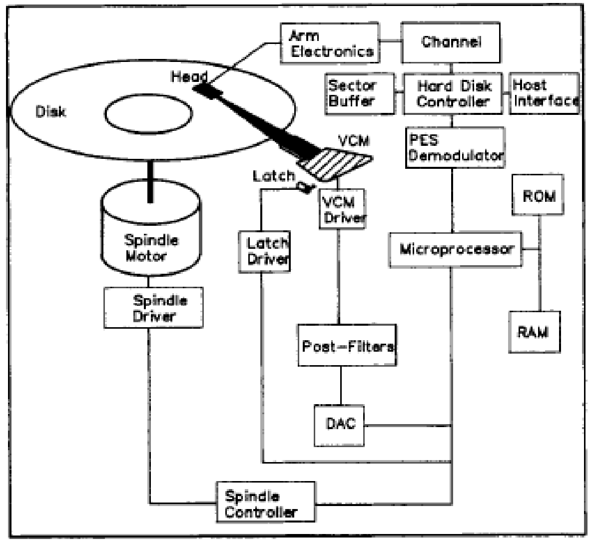
- System-level power management
  - Saves power of subsystems (devices)
- Device is:
  - Busy if there are requests
  - Idle otherwise
- Changing power states incur some overhead
  - T_{sd} => shutdown delay
  - T_{wu} => wake-up delay

Power State & Transitions



- Assumption
  - Only one device & only one stream of requests

HDD Organization

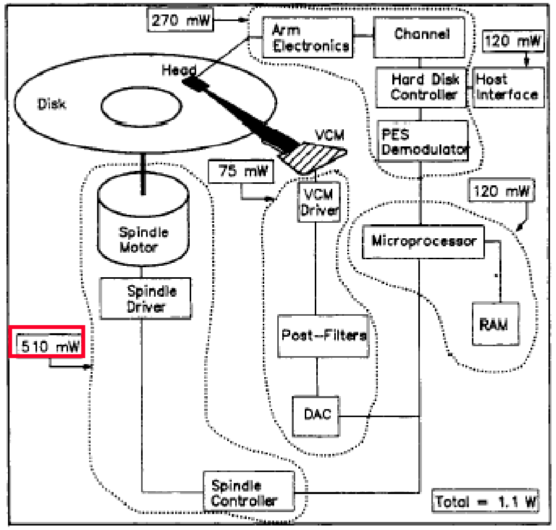


HDD Power States [Harris *et al.* 1995]

Form-Factor	3.5-in	2.5-in	1.8-in	1.3-in
Key Parameters				
Capacity (MB)	720	340	85	40
Weight (g)	500	180	75	28
Height (mm)	20	12.7	10.5	10.5
Disk Diameter (mm)	95	65	48	34
Number of Disks	2	2	2	2
RPM	4500	3800	4500	5400
Power Requirements				
Start-Up (W)	8.0	4.7	3.0	2.2
Seek (W)	4.5	2.0	1.5	1.7
Read/Write (W)	4.2	2.8	1.725	1.6
Idle (W)	2.5	1.35	0.75	1.0
Standby (W)	0.8	0.4	0.035	0.5
Sleep (W)	0.1	0.2	0.02	0.015

Trade-off between recovery time (to read/write) and energy saving

Idle Mode Power Consumption



A 2.5-in HDD
With 2 disks
At 3600 RPM

Operation Modes

- Startup: Spindle accelerated from rest to rated speed.
- Seek: VCM is actively moved to a new position.
- Read/Write: Data is transferred between media and head while on a track.
- Idle: Spindle and actuator are under normal control. No read/write of data. Only PES information is read by the demodulation circuit.
- Standby: Spindle is at rest, heads are parked, and interface can receive commands from host.
- Sleep: In addition to “Standby” condition the interface is powered off and a single logic line is active to sense a reset signal from host.

PES = Position Error Signal
VCM = Voice Coil Motor

Power Management Policies

- Policy classification
  - Time-out
    - TO (Time-Out)
    - ATO (Adaptive Time-Out)
    - DDT (Device-Dependent Time-out)
  - Predictive
    - L-shape
    - LT (adaptive Learning Tree)
    - EA (Exponential Average)
  - Stochastic
    - DM (Discrete-time Markov process)

Oracle Power Manager

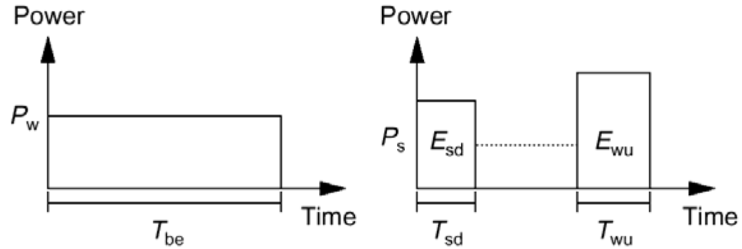
- Perfect information about future request
  - Shut down the device immediately after the device becomes idle
  - Lower bound of power consumption
- Does not exist in practice
- Reference point for comparison

Time Out Policies

- τ => time-out value
  - Fixed τ
  - Adjustable τ
- T_{be} => break-even time

Break-Even Time: T_{be}

- T_{be} makes the energy consumption in both cases equal



$T_o = T_{sd} + T_{wu}$ **Transition time**

$E_o = E_{sd} + E_{wu}$ **Transition energy**

$T_{ms} = T_{be} - T_o$ **Minimum sleeping time**

T_{be} & T_{ms}

Pw x Tbe

Eo + Ps x (Tbe – To)

$$P_w \times T_{be} = E_o + P_s \times (T_{be} - T_o)$$

$$T_{be} = \frac{(E_o - P_s \times T_o)}{(P_w - P_s)}$$

Time-Out Policy

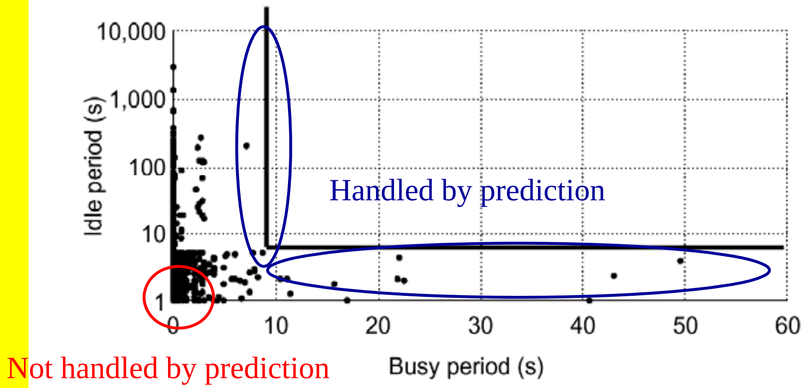
- Assumption
  - After a device is idle for τ, it will remain idle for at least T_{be}
- Theorem:
  - If τ = T_{be}, energy consumption of the policy will be at most twice the energy of an ideal oracle policy
    - 2-competitive algorithm

Adaptive Time-Out

- Time-out value adjusted at run time
  - by the spin-up delay / the previous idle period
  - If the ratio is too small
    - Decrease time-out value τ
      - Wasting less energy waiting for time-out
  - If the ratio is too large (i.e., a bump occurs)
    - Increase time-out value τ
      - Reducing inappropriate spin-down's

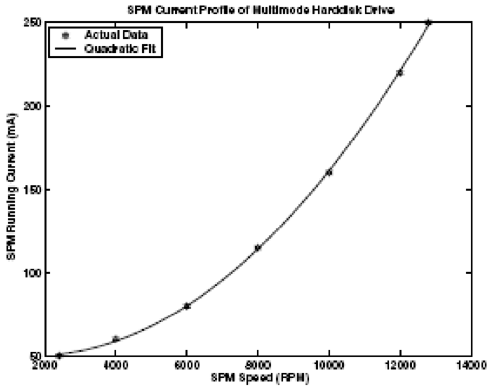
L-shape Predictive

- Based on an L-shape pattern bet. busy and idle periods:
  - Short busy period -> long idle period
  - Long busy period -> short idle period
- Shut down if busy for a short period!



DRPM [Gurumurthi *et al.* ISCA03]

- Dynamic Rotations Per Minute
  - Dynamically modulate the speed at which the disk spins
    - Quadratic power savings by decreasing the spinning speed



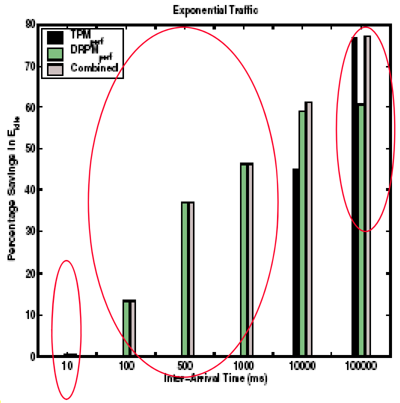
Server Workload

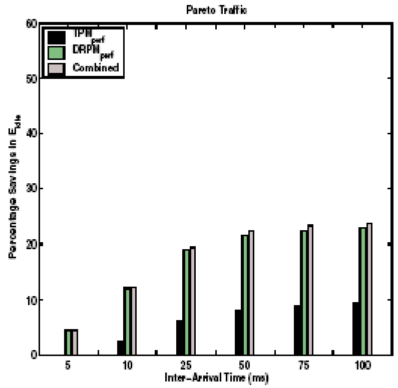
DRPM over Traditional PM

- Adapt better for short idle intervals
  - Less delay for a spin up
  - Less energy for short idle intervals
- Better power-performance tradeoff
  - Continuous vs. Discrete (i.e., 0 & MAX)

Energy Saving by DRPM

Bursty requests

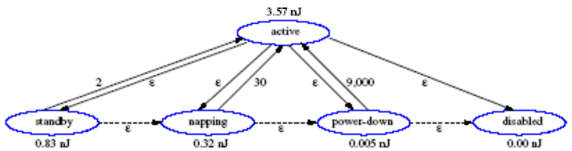




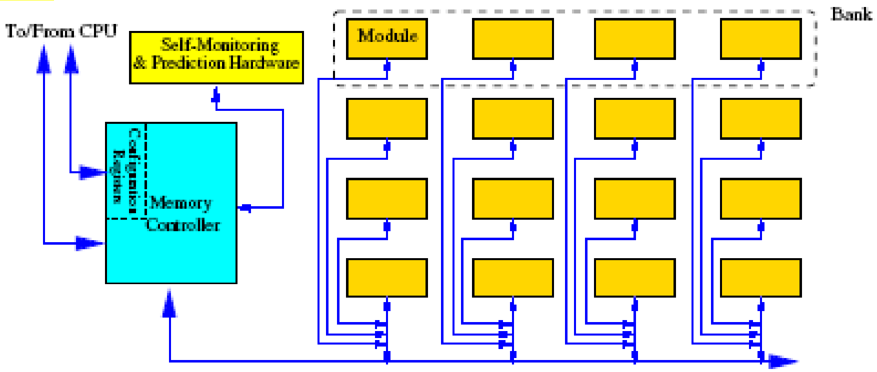
TPMperf & DRPMperf: based on an idle time prediction oracle
Combined = Choose TPM or DRPM using an oracle

DRAM Energy Management [Delaluz *et al.* HPCA01]

- Similar to HDD Management Problem
- Operating Modes:
  - Active
  - Standby
  - Napping
  - Power-down
  - Disabled
- Each mode differs in:
  - Energy consumption
  - Resynchronization time
  - The lower the energy consumption, the higher the resynchronization time



Memory System Architecture



System Support for Mode Setting

- Self-Monitored Approach
  - Self-Monitoring and Prediction Hardware block
    - Monitors ongoing memory transactions
    - Estimate the time until the next access to a memory bank
- Software-Directed Approach
  - Configuration Registers in the memory controller
    - Memory-mapped I/O
    - Storing predefined values to these registers initiates mode changes
  - Should be O/S's job unless a single program environment

Self-Monitored Approach

- Adaptive Threshold Predictor
  - High hardware costs (in DRAM setting)
- Constant Threshold Predictor
- History-based Predictor
  - Overcome:
    - the problem of gradual mode changes
    - the problem of resynchronization cost
  - Estimate the inter-access times

ACPI

- Advanced Configuration and Power Interface
  - Promoted by Intel, Microsoft, Toshiba
  - Interface Spec. between H/W and S/W for OS-directed system-level power management.

ACPI Interface & PC Platform

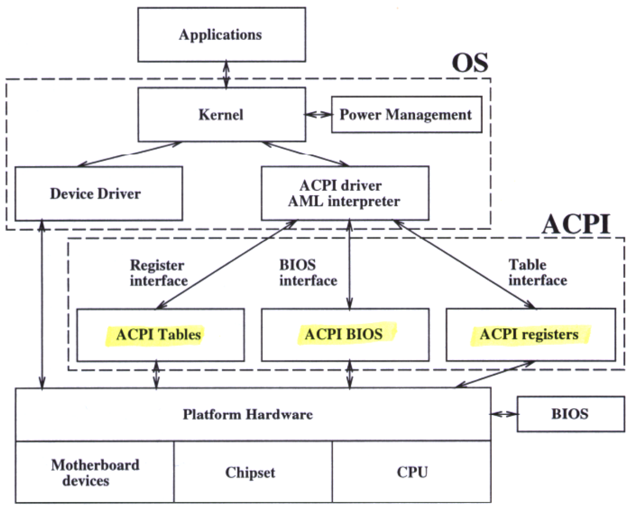


Figure 6.6. ACPI interface and PC platform

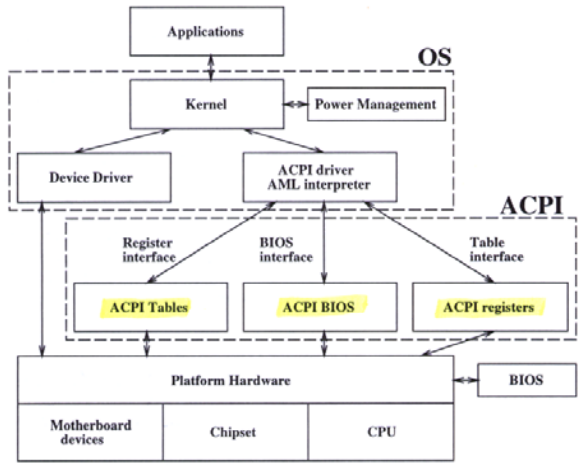


Figure 6.6. ACPI interface and PC platform

ACPI Components

- ACPI DOES NOT specify how to implement hardware devices or the power management policy in OS.
- ACPI Tables
  - Interface to hardware
- ACPI BIOS
  - Portion of firmware compatible with ACPI
- ACPI Registers
  - Constrained part of the hardware interface

ACPI States Definitions

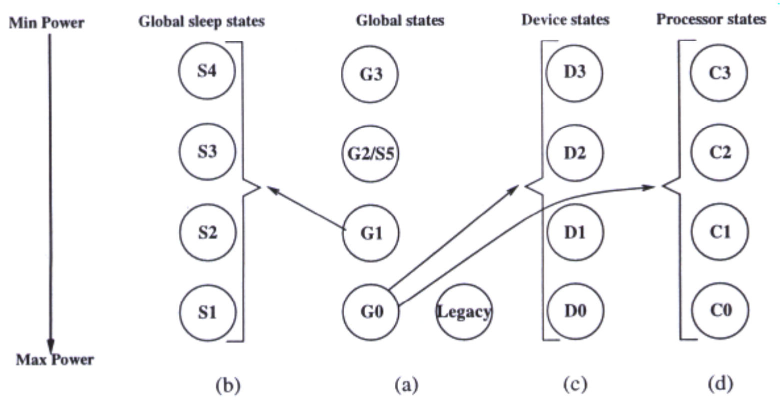


Figure 6.7. State definitions for ACPI

ACPI State Definitions

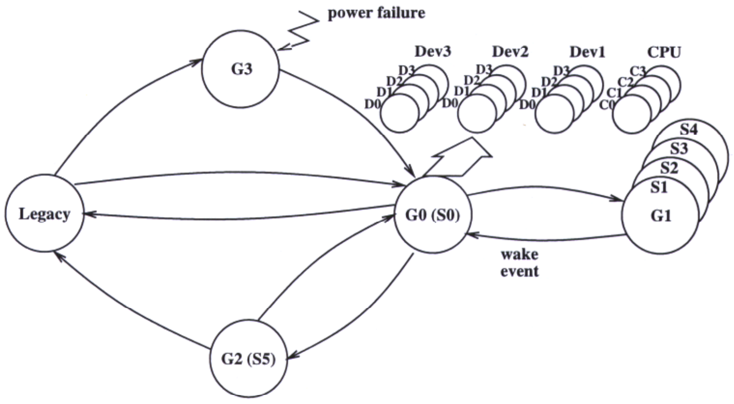
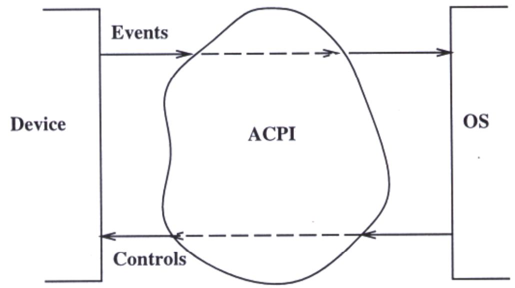


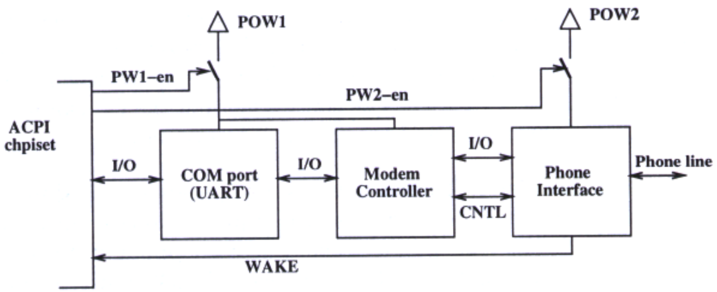
Figure 6.8. Global and power states and substates

Device Power Management



- Identification of device power capabilities
- Setting device power states
- Getting device power states
- Enabling device-controlled wakeup

Example: Integrated Modem



- D0
- D1: MC & phone interface in low-power mode
- D2:
- D3: MC is off, phone interface powered by phone line or off

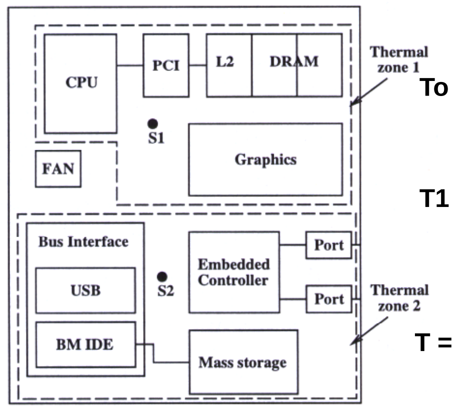
Battery Management

- Must confirm either Smart Battery subsystem interface or a control method battery (CMBatt) interface
- CMBatt reports
  - The designed capacity
  - The latest full-charged capacity
  - The current remaining capacity
  - Warning, Low & Critical status messages

Thermal Management

- Based on thermal zones
  - Using temperature events
- Two types of cooling
  - Active cooling
    - Exploits cooling devices (e.g., fans)
    - Increases power to reduce heat
  - Passive cooling
    - Reduces the power-consuming activities
    - Reduces power to decrease temperature

Example: Thermal Zones



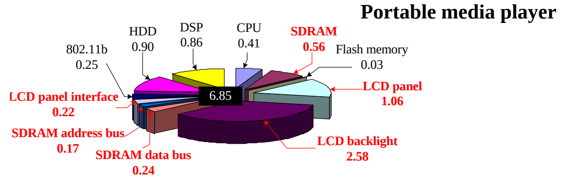
To = 50°C
passive cooling activated
CPU clock slowed down by 2

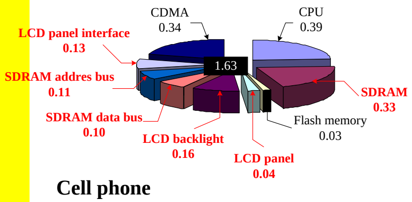
T1 = 70°C
active cooling activated
fan is started

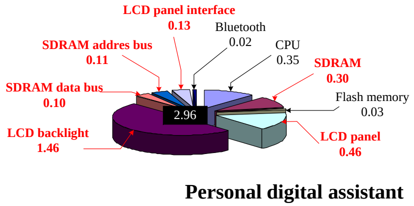
T = 60°C, fan is started

Figure 6.11. Thermal zones in a PC board

System-level Power Breakdowns [Shim 06]







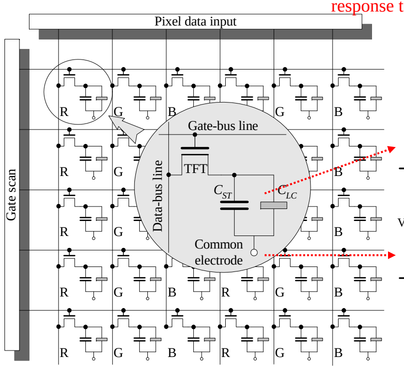
Power consumption for running a streaming video application (W)

Display Power Management

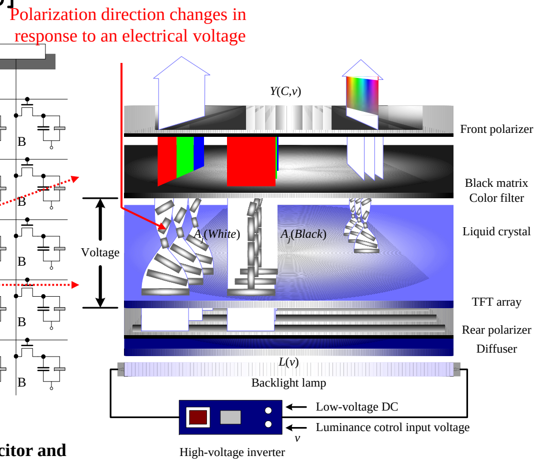
- Idle-based display off
- Brightness control based on user request
- Active mode power management
  - Zone-based on & off control
  - Dynamic Luminance Scaling (DLS)

Liquid Crystal Displays [Shim 06]

- TFT LCD panel [TFT96]



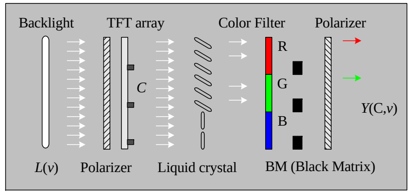
TFT-array substrates (C_{ST}: a storage capacitor and C_{LC}: an equivalent capacitor for a liquid crystal cell)



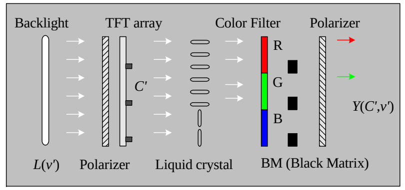
Structure of a transmissive LCD panel

Overview of DLS [Shim 06]

- Principle of DLS (Dynamic Luminance Scaling)
  - Backlight dimming for power saving
    - $L(v) \Rightarrow L(v')$ such that $L(v') < L(v)$
  - Restoring brightness/contrast by appropriate image compensations
    - $C \Rightarrow C'$ to maintain the same intensity perceived by human eyes



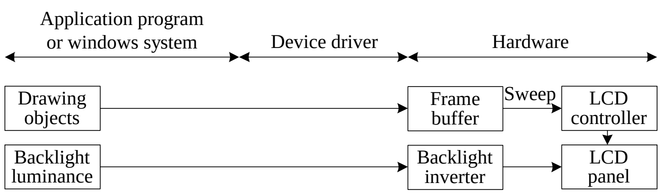
Original



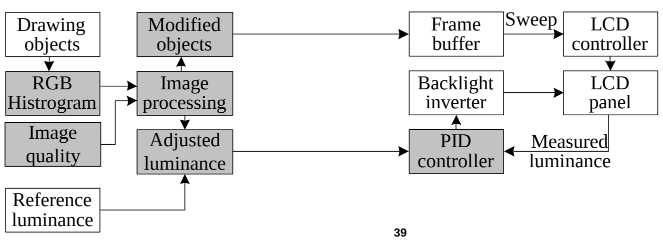
Backlight dimming with appropriate image compensation

One Implementation

Conventional



Application embedded or windows system embedded



Experimental Results

- MPEG-4 decoder (application)

Power reduction/Relative backlight luminance

Original DLS (D_i=0.03)

What is Low Power Communication?

- Recently, mobile communication device are ubiquitous
  - Mobile phone, PDA….
- Mobile devices are operated by battery
- Low power communication
  - Reduce power consumption in communications
  - Longer communication time is possible with limited battery
  - It includes
    - Make low power communication devices
    - Make more efficient dynamic power management scheme
    - Make power aware routing protocols

Dynamic Power Management of Mobile device

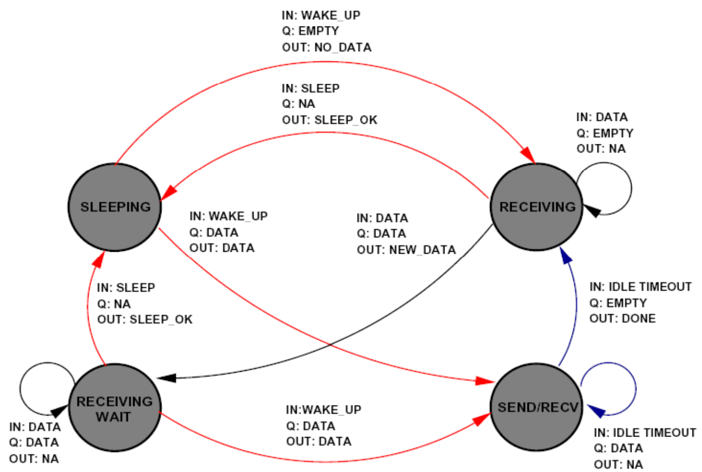
- Motivation
  - Wireless interface cards consumes energy continuously
    - 10 to 50% of energy budget for mobile device
- Different level of power state
  - When idle, enter lower power state
- If a state is in a suspend state
  - Cannot know if some other host has data to send to it
  - External events need to wake up
- The key of balancing power saving & delay
  - Knowing when to suspend & wake up
    - Role for application specific information for guidance
- Example
  - Power Management Techniques for Mobile Communication [Kravets & Krishnan, '98]

Kravets & Krishnan '98

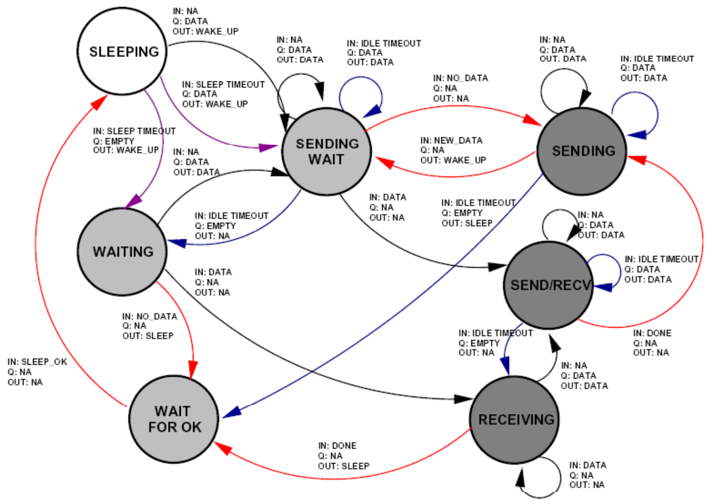
- A transport level protocol for managing the suspend/resume cycle of the mobile host's communication device
- Idle communication periods are detected by
  - Timeout periods
  - Application Signalling

Protocol State Diagram: Base Station

- Mobile Host: Master
- Base Station: Slave



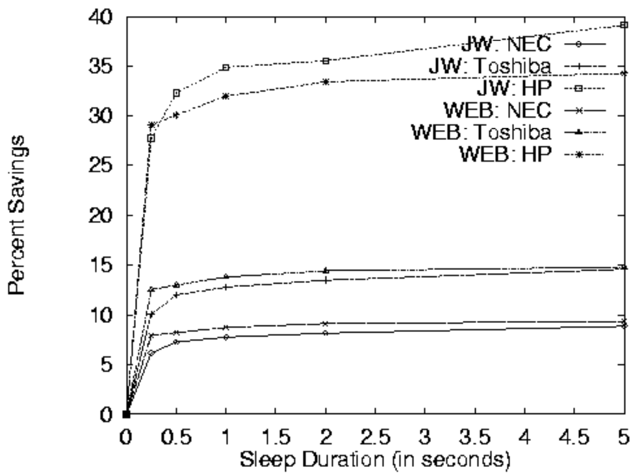
Protocol State Diagram: Mobile Host



Experimental Results

- Saving for the three types of machines

Machine	Power Requirements Idle w/o WaveLAN Card	Power Requirements Idle w/ WaveLAN Card
NEC Versa 6320	14W	15.5W
Toshiba Libretto 60	7W	8.5W
HP Palmtop PC 320LX	1.2W	2.7W



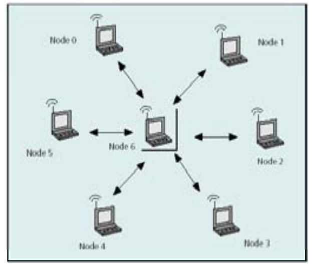
Ad Hoc Network

- What is wireless Ad Hoc Networks?
  - Infrastructure-less networks (no fixed routers)
  - All nodes capable of movement
  - Links appear and disappear dynamically
  - Special constraints:
    - Limited bandwidth
    - Limited power: need power awareness
    - High error rates
  - Protocol needed to create and maintain routes
  - Conventional routing protocols not applicable

Power-Aware Ad Hoc Network

- Power-Aware Ad Hoc Networks
  - Most mobile hosts of ad hoc network are based on batteries.
  - Power-aware ad hoc network increases the battery life time and system life time.
- Power reduction can be obtained in several layers:
  - Physical layer
    - Adjust transmission power [Ramanathan *et al*, '00]
  - Data link layer
    - Sleep mode operation [Singh *et al*, CCR'98]
  - Network layer
    - Maximize system life time [Singh *et al*, MobiCom'98] [Toh, '01]
  - Transport layer
    - Reduce protocol processing and software overhead [Agrawal *et al*, '01]

Metrics in Ad Hoc Routing Protocols



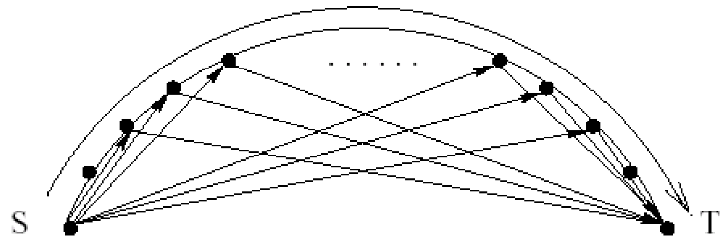
- Shortest-hop
  - Some nodes die relatively early
    - E.g., the middle node (node 6)
- Message/time overhead
  - Routing protocol message overhead
  - Misguided in the long-term
    - Control packet: 5 ~ 10 %
    - Data packet: 90 ~ 95%
      - Amdahl's Law
- Link quality and/or location stability
  - Stability → less table update
  - Orthogonal to the goal of power-awareness
    - Can be used in conjunction

Power-Aware Metrics [Singh *et al*, MobiCom'98]

- Minimize energy consumed / packet
  - Under the light loads, similar to shortest hop
  - Under the heavy contentions, avoid congested areas
- Maximize time to network partition
  - Load balancing among nodes
- Minimize variance in node power levels
- Minimize cost / packet
  - Node cost == node's reluctance to forward packets
  - Node 6 case avoided

Online Max-Min Routing [Li *et al.*, 2001]

- Max Min Path
  - The path with the maximal minimal fraction of remaining power after the message is transmitted
  - (–) Poor performance (i.e., delay increases)
  - (–) Large power consumption



- Trade off between
  - Minimizing the end-to-end delay
  - Maximizing the minimal residual power

Max-Min zP_{min} Path

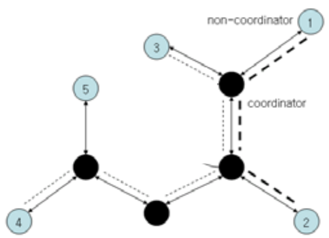
- Two extreme solutions
  - A path with minimal power consumption P_{min}
  - A path that the maximizes the minimal residual power in the network
- Max-Min zP_{min} path
  - Relax the minimal power consumption to $z \times P_{min}$

Zone-Based Routing

- Very expensive to implement Max-Min zPmin
- Solution: Zone-based routing
  - Group together all the nodes that are in geographic proximity as a zone, treat the zone as an entity in the network
  - Require 3 subtasks
    - Zone Power Estimation
    - Global Path Selection
    - Local Path Selection

Coordination in Power Saving [Chen *et al.* '01]

- In ad-hoc networks, not all nodes participate in routing
- By using some nodes as a coordinator
  - Extend life time
  - Reduce overall energy-consumption
  - Preserve network capacity



References

- R. Ramanathan, and R. Rosales-Hain, "Topology control of multihop wireless networks using transmit power adjustment", Proc of INFOCOM, 2000
- S. Shingh, and C.S. Raghavendra, "PAMAS-Power aware multi-access protocol with signaling for ad hoc networks", ACM CCR, 1998
- S. Shingh, M. Woo, and C.S. Raghavendra, "Power-aware routing in mobile ad hoc networks", Proc of MobiCom, 1998
- C.-K. Toh, "Maximum battery life routing to support ubiquitous mobile computing in wireless ad hoc networks", IEEE Communication Magazine, 2001
- Q. Li, J. Aslam, and D. Rus, "Online Power-aware Routing in Wireless Ad-hoc Networks", Proc of SIGMOBILE, 2001
- J. Chang, and L. Tassiulas, "Energy conserving routing in wireless ad-hoc networks", Proc of INFOCOM, 2000
- K. Scott, and N. Bambos, "Routing and Channel Assignment for Low Power Transmission in PCS", Proc of ICUPC, 1996
- Y. Xu, J. Heidemann, and D. Estrin, "Geography-informed Energy Conservation for Ad hoc Routing", Proc of MobiCom, 2001
- B. Chen, K. Jamieson, H. Balakrishnan, and R. Morris, "SPAN:An Energy-Efficient Coordination Algorithm for Topology Maintenance in Ad Hoc Wireless Networks", Proc of SIGMOBILE, 2001
- R. Kravets, and P. Krishnan, "Power Management Techniques for Mobile Commnunication",Proc of MobiCom, 1998

References

- Y. Lu and G. D. Micheli, “Comparing System-level Power Management Policies”, IEEE Design and Test, Mar./Apr. 2001