Module III

Data storage and communication

Storage Media Power Characteristics

Hard Disks

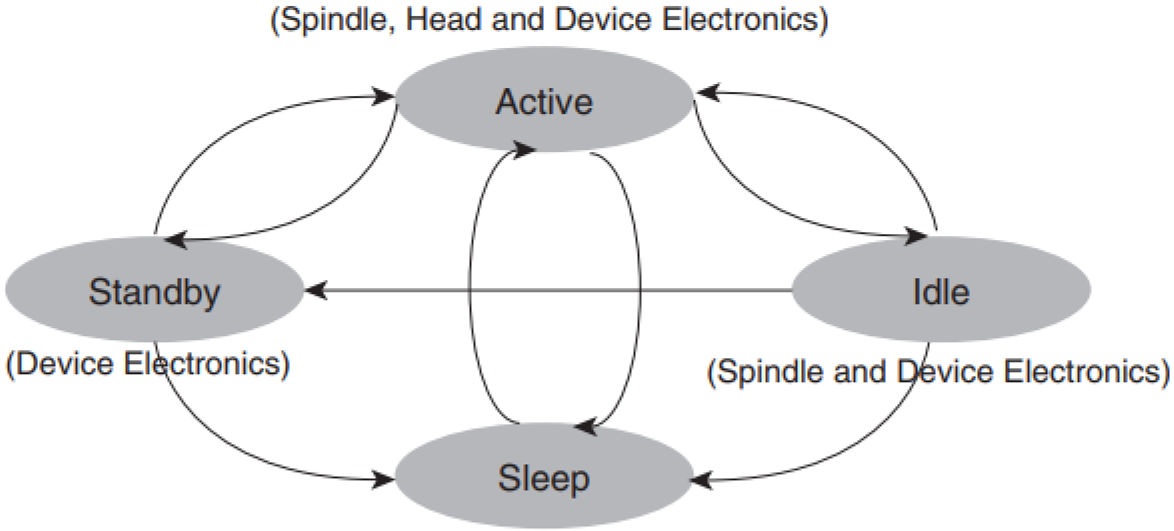
- Hard disks are the most common non volatile storage media.
- A hard disk drive (HDD) contains disk platters on a rotating spindle and read-write heads floating above the platters. The read-write heads encode data magnetically.
- Power consumption of a hard disk in different operational (power) states can differ considerably, and energy management techniques aim to effectively exploit this feature.
- Power consumption of magnetic hard disks is a function of its rotational speed and the data access rate.
- Most power is consumed by the rotating spindle, followed by the head assembly that moves along the platters to the requested sectors or logical block addresses (LBAs) and the buffers used for queuing requests and requested data.
- The total power consumption (P_{total}) of a hard disk is the sum of the power consumed by the spindle motor ($P_{spindle}$), the power consumed by the head movement (P_{head}) and the power consumed by other components (Pother) which is relatively small, has less variation and includes the power consumed by buffers (P_{buffer}).

$$P_{total} = P_{spindle} + P_{head} + P_{other}$$

- The power consumed by the spindle motor ($P_{spindle}$) is directly proportional to the square of its angular velocity (ω), that is, $P_{spindle} \propto \omega 2$.
- A hard disk is either rotating at its full speed or not rotating at all; therefore, $P_{spindle}$ is power consumed at full speed ($P_{fullspeed}$) or zero.
- The power consumed by the head movement (P_{head}) is dictated by the disk access pattern. Random access causes more head movements than sequential accesses, and thus leads to higher head power consumption.

- In addition, when disks are placed in large disk array–based systems, the array controllers and enclosures consume an additional 1–2 W of power per disk.
- These states are governed by the power states of the three main subcomponents – the spindle motor, head assembly and device electronics.
- In the active state, all the three components are powered on and input–output (I/O) is serviced.
- The head assembly and the buffer are both on, and the spindle is rotating at its full speed.
- The power consumption in active mode is defined as follows: $P_{active} = P_{fullspeed} + P_{head} + P_{other}$.
- In this state, power consumption varies based on the degree of head movement.
- If read-write requests are sequential in nature, head movement is relatively minimal and hence power consumption is mainly determined by the spindle’s rotational speed.
- In the idle state, only the spindle motor and device electronics are on and the head typically does not consume any power in this state.
- The head assembly and the buffer are both on, and the spindle is rotating at full speed.
- However, in this state no requests are actually being serviced, and thus $P_{head} \approx 0$.
- Power consumed is mainly determined by the spindle’s rotational speed, $P_{idle} = P_{fullspeed} + P_{other}$.
- Transiting between the idle and active states is instantaneous. In the standby state, only the device electronics are on. In this state, the commands received on the interface are queued for servicing.
- The spindle is at rest ($P_{spindle} = 0$), but the buffers are on which facilitates queuing of new requests, and the disk is still able to respond to diagnostic queries by the controlling system.
- Power consumed in this state, $P_{standby} = P_{other}$, is very little compared to the active or idle state.
- The main disadvantage, however, is the transitioning from standby to idle or active takes a long time, typically about 8–10 seconds to get the spindle running at its specified speed.

States of mechanical hard disk drives


(Spindle, Head and Device Electronics)
Active
Standby
Idle
(Device Electronics)
(Spindle and Device Electronics)
Sleep

- In the sleep state, all components are off and hence any requests will result in IO timeouts or errors.
- Important to note here is the difference in transitioning times between different states and the power consumed for the transition.
- Switching from the standby to active state involves restarting the spindle motor and hence is on the order of seconds, typically about 8–10 seconds.
- Further, peak current during the transition is significantly higher than during steady-state usage.
- State transitioning (a process of transitioning a hard disk to different operational states based on its access pattern or idle period) is a commonly used technique for storage energy management.
- To increase the idle periods, which are required for state transitioning to work well, disk accesses are minimized by caching.
- Hard disks can be accessed by the host system via different types of bus.
- Based on the bus type, hard disks can be categorized into advanced technology attachment (ATA), also called integrated drive electronics (IDE), serial advanced technology attachment (SATA), small computer system interface (SCSI), serial attached SCSI (SAS) and fibre channel (FC).
- Each of these hard drive categories has different power characteristics as they have different hardware and operational profiles, such as their type of magnetic material used, number of platters, amount of buffering and rotational speed.
- For instance, SCSI, SAS and FC drives are typically targeted at enterprise computing and hence have high rotational speeds leading to higher active or idle power consumptions as compared to ATA, IDE and SATA drives.

Magnetic Tapes

- Magnetic tape is another medium for data storage.
- It is made of a thin magnetic coating on a long, narrow strip of plastic.
- To store data on magnetic tapes, a tape drive uses a motor to wind the magnetic tape from one end to another and through tape heads to read, write and erase the data.
- Magnetic tape is still a major data storage medium especially in some industries such as radio and TV broadcasting because of its very low storage-per-bit cost compared to that of hard disks.
- Although its areal density is lower than the density of disks, the capacity of tape is still very high (i.e. similar to disks) due to its larger amount of surface area compared to disk surface.
- The large capacity and cost advantage makes magnetic tapes a viable alternative, particularly for backup and archiving.
- Further, the average energy consumption of tape drive is much less compared to a hard drive as it consumes power only when it reads or writes.
- Power consumption in the idle or retention state is zero.
- The cost of maintaining a tape library, which can be amortized over thousands of tape cartridges, is negligible.

Solid-State Drives (SSDs)

- Recently, SSDs have gained increased adoption due to their better performance potential and higher power efficiency.
- Unlike electromechanical HDDs which contain spinning disks and movable heads, a SSD uses solid-state memory (i.e. nonvolatile memory) to store persistent data and has no moving parts.
- Hence, compared to HDDs, SSDs have lower access time and power consumption.
- Today’s SSDs use nonvolatile not AND (NAND) flash memory.
- There are two types of flash memory: multilevel cell (MLC) and single-level cell (SLC) flash memory.
- SLCs store a single bit in a single memory cell, whereas MLCs store multiple bits in a single cell by allowing each cell to store multiple electrical states.
- For example, a four-level MLC stores 2 bits in a single memory cell.
- Typically, MLCs are cheaper and have greater storage density, but are slower than SLCs.
- SSD has rapidly increased in popularity as the primary data storage medium for mobile devices, such as phones, digital cameras and sensor devices, notebook computers and tablets.
- Its features include small size, low weight, low power consumption, high shock resistance and fast read performance.
- SSD has also started to penetrate the markets from laptops and PCs to enterprise-class server domains.
- Enterprise-class SSDs provide unique opportunities to boost I/O-intensive applications’ performance in data centres.

- Compared to hard disks, SSDs are very attractive for the high-end servers of data centres due to their faster read performance, lower cooling cost and higher power savings.
- SSDs have the lowest power consumption rate of active devices when compared with both dynamic random-access memory (DRAM) and hard disks.
- For example, the power consumption for a 128 GB SSD is about 2 W, that of a 1 GB DRAM dual in-line memory module (DIMM) module is 5 W, and for a 15 000 RPM, 300 GB (7200 RPM, 750 GB) hard drive, it is around 17.5 W (12.6 W).
- This efficiency can mainly be attributed to the nonmechanical nature of these devices.
- Data in SSD is read or written in units of pages.
- The read time for a page in SSD is about 20 times faster than a page read from hard disks.
- Unlike hard disks, there are no differences in terms of time between random reads and sequential reads.
- SSD is best suited for caching data between memory and hard disks since the hard disks still hold the advantage of cost and storage capacity.
- However, the relatively low write performance and longevity problems of SSD require new and innovative solutions to fully incorporate SSDs into the high-end servers of data centres.

- Unlike conventional disks, where read and write operations exhibit symmetric speed, in non-enterprise SSDs, the write operation is substantially slower than the read operation.
- This asymmetry arises because flash memory does not allow overwrite and write operations in a flash memory must be preceded by an erase operation to a block.
- Typically a block spans 64–128 pages, and live pages from a block need to be moved to new pages before the erase can be done.
- Furthermore, a flash memory block can be erased only a limited number of times.
- The slow write performance of non-enterprise SSDs and the wear-out issues are the two major concerns.
- Existing approaches to overcome these problems through modified flash translation layer (FTL) are effective for sequential write patterns; however, random write patterns are still very challenging.

Energy Management Techniques for Hard Disks

- To reduce the energy consumption of hard disks, different techniques and methodologies are being adopted.
- Three commonly used energy management techniques for hard disks are state transitioning, caching and dynamic RPM.

State Transitioning

- Given that in a hard disk, the spindle motor consumes most of the power, state transitioning techniques try to turn off the spindle motor or keep it in standby mode during idle periods.
- The disk transitions to standby or off mode if there is no request to be served.
- If the disk idle period is not long enough, the time overhead of spin down and spin up can affect the disk response time significantly.
- Moreover, the power consumed by transitioning itself might exceed the power saving gained from a low-power state during the short idle period.
- If the disk is already idled for the threshold time, it transitions to standby mode.
- If it stays in standby mode for another threshold of time without requests, it can further transition to off mode.
- In this approach, the historical information is used to predict the future access pattern.
- Different variations on access prediction are based on historical information.
- Most on-going research and development in state transitioning revolve around idle period prediction and minimizing the performance impact of these transitions on disk responsiveness (as the transition time is usually around 8–10 seconds).
- Some state-transitioning techniques provide a performance guarantee.

Caching

- In order to speed up access for both read and write requests, enterprise storage solutions typically have huge amounts of cache in conjunction with regular disks.
- Further, to make use of the cache to aid in disk power management, various techniques are recommended.
- These cache management techniques or algorithms aim to minimize disk power usage, either by minimizing disk access or by increasing the length of idle periods.
- One could in effect use huge caches to increase the idle periods of disks and in doing so can help more disks to transition to the sleep state, thereby improving energy efficiency.
- The cache management algorithms partition-aware least recently used (PALRU) (Zhu et al., 2004) and partition-based LRU (PBLRU) (Zhu, Shankar, and Zhou, 2004) are centred on this idea.
- PALRU classifies all disks based on access patterns into two classes – priority (disks with fewer cold misses and longer idle times) and regular – and maintains two separate LRU queues.
- At the time of an eviction decision, first the regular queue elements are chosen as victims.
- If the regular queue is empty, the algorithm chooses elements from the priority queue.

- PBLRU, however, differentiates between disks by dynamically varying the number of allocated cache blocks per disk.
- It divides the cache into multiple partitions (one per disk) and adjusts the size of these partitions periodically based on workload characteristics.
- The simulation results with online transaction processing (OLTP) traces show that PALRU consumes 14–16% less energy and PBLRU consumes 11–13% less energy than traditional LRU.
- On the other hand, for the Cello96 trace (file system trace), PALRU saves less than 1% energy over LRU whilst PBLRU is 7.6–7.7% more energy efficient than LRU.
- Since write requests in enterprise storage devices almost never get written directly to target disks (they are cached instead), another technique is to use write offloading as a mechanism to conserve disk power usage (for details, see Narayanan, Donnelly, and Rowstron, 2008).
- Write offloading facilitates complete spin downs of volumes periodically, thereby aiding in significant power savings.
- By using write offloading, about 45–60% energy savings can be achieved in write-dominated application environments.

Dynamic RPM

- Dynamic RPM in which the rotation speed of a hard disk is varied based on workload is another technique for hard disk energy management.
- It assumes availability of multispeed hard disks, and power consumption increases with the speed of rotation.
- In dynamic RPM, the rotational speed of the disk is altered based on the desired response time of disks and the performance requirement.
- A fast response time that is greater than the specified or expected threshold is a waste of performance.
- The idea here is to limit this wastage of performance by switching the rotational velocity of the disk to a lower value that still yields acceptable performance.
- Practical implementation of this approach is limited by the feasibility of developing a single disk that can change speeds in a cost-effective manner, but simulation results reveal that a dynamic RPM scheme can yield a power savings of up to 60%.

Objectives of Green Network Protocols

- In empowering the network with energy awareness and efficiency ability, it is necessary to understand protocol overhead in terms of mandatory fields in packet headers and control packets currently used to manage transmissions.
- The objective of this section is therefore to gain an appreciation for the way in which protocols may be optimized whilst application QoS is maintained, and to understand mandatory content carried within protocol headers.
- This leads to identification of ways in which protocols may be optimized such that their degree of reliability is maintained through reducing the number of bits associated with each, the cost of which to transmit may be incurred during any transaction, and hence energy efficiency improved.

Energy-Optimizing Protocol Design

- Network efficiency can be enhanced by the design of protocols used.
- Reducing the number of bits associated with a transmission and minimizing network load will optimize communication efficiency.
- Where fewer bits are transmitted, less processing operation will be required at nodes, fewer finite power resources consumed during transmission and less carbon emitted, along with less congestion in the network, fewer retransmissions and an overall more optimized process.
- From the point of view of network protocols, the number of bits involved can be reduced by

  (i) minimizing the number of overhead packets per protocol,

  (ii) minimizing the number of mandatory bits per protocol,

  (iii) minimizing retransmission attempts and

  (iv) maximizing the number of successful data packets sent.

Objective 1: minimizing the number of overhead packets O per protocol

$$\text{minimize } O\ n{\in}N{\ast}(i,j\ ;G)$$

- where O represents the number of overhead packets associated with a transmission, pushed from each node n where $n \in N{\ast}(i,j\ ;G)$ is the set of all nodes traversed across sublinks on path (i,j) between source and destination devices in network G.
- ‘Overhead’ in this case refers to control and management packets transmitted across the network in support of the protocol design.
- In the case of the Ad hoc On-Demand Distance Vector (AODV) protocol (Perkins, Belding-Royer, and Das, 2003), for example, a broadcast packet is sent when a connection between nodes which wish to communicate is needed.
- Intermediary nodes forward the message towards the destination; when the message is received at a node with a route to the destination, it communicates this detail with the source node, which subsequently begins to use the route.
- In minimizing the number of overhead packets used to support protocol operation, optimization in power requirements can be achieved.

Objective 2: minimizing the number of mandatory M bits per protocol

$$\text{minimize } M \; n \in N\ast(i,j\ ;G)$$

- where M is the number of mandatory bits associated with a protocol for packets pushed from each node n using the protocol where $n \in N\ast(i,j\ ;G)$ is the set of all nodes traversed across sublinks on path (i,j) between source and destination devices in network G.
- Mandatory bits include those transmitted alongside application data in packets encapsulated at each stack layer.
- In minimizing the number of mandatory bits associated with a protocol, fewer resources will be required to support packet transportation, leading to an overall more optimized, and subsequently efficient, communication.
- In parallel with optimizing the number of bits associated with a network transaction, sufficient detail and capability should be maintained such that operational performance is achieved.

Objective 3: minimizing retransmission attempts R

$$\text{minimize } R \; n{\in}N{*}(i,j \;;G)$$

- where R is the number of retransmission attempts associated with traffic pushed from each node n where $n \in N{*}(i,j \;;G)$ is the set of all nodes traversed across sublinks on path (i,j) between source and destination devices in network G.
- ‘Retransmissions’ refer to data packets sent more than once through the network when reliable protocol mechanisms have been applied in the instance that application packets have been lost or received incorrectly.
- One or more retransmissions may be sent in response.

Objective 4: maximizing the number of successful data packet sends S

$$\text{maximize } S \ n{\in}N{\ast}(i,j\ ;G)$$

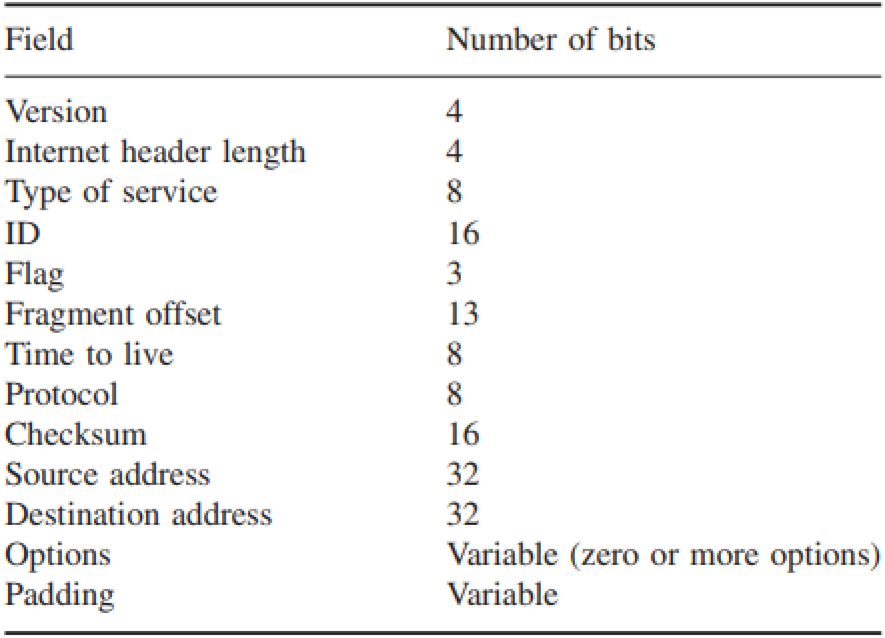
where S is the number of packets sent successfully from each node n where n $\in$ $N{\ast}(i,j\ ;G)$ is the set of all nodes traversed across sublinks on the end-to-end path (i,j) within network G.

Bit Costs Associated with Network Communication Protocols

- Reducing the overhead costs associated with protocols will improve their energy efficiency.
- Minimum costs of network protocols and typical overhead volumes associated with packets sent using each protocol are therefore explored in this section to highlight the network resources required for protocols to operate and lead to optimization of their design through identification of inefficiencies.

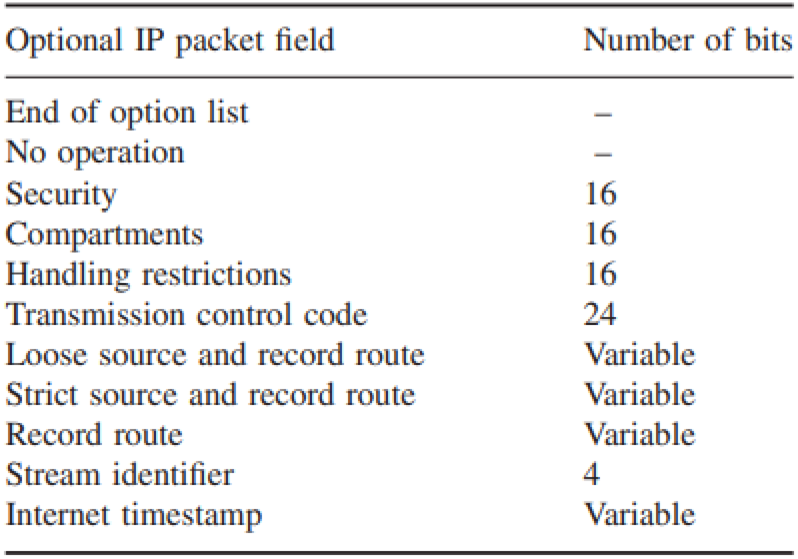
Internet Protocol v4 (IP) Cost

IPv4 packet header format



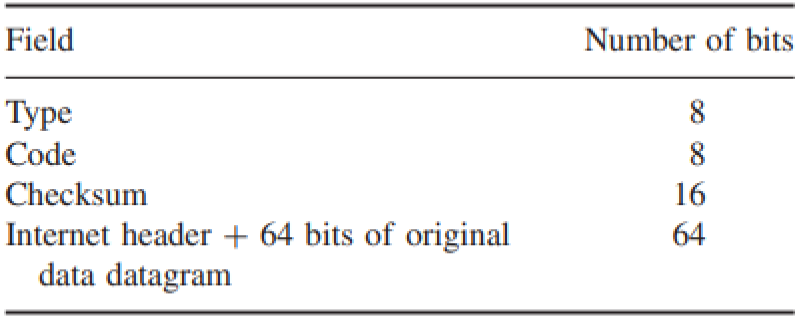
Field	Number of bits
Version	4
Internet header length	4
Type of service	8
ID	16
Flag	3
Fragment offset	13
Time to live	8
Protocol	8
Checksum	16
Source address	32
Destination address	32
Options	Variable (zero or more options)
Padding	Variable

Optional IPv4 packet header fields



Optional IP packet field	Number of bits
End of option list	–
No operation	–
Security	16
Compartments	16
Handling restrictions	16
Transmission control code	24
Loose source and record route	Variable
Strict source and record route	Variable
Record route	Variable
Stream identifier	4
Internet timestamp	Variable

ICMPv4 destination unreachable message packet header format

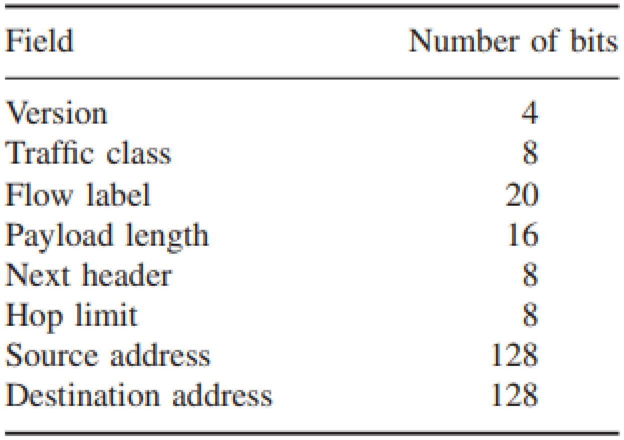


Field	Number of bits
Type	8
Code	8
Checksum	16
Internet header + 64 bits of original data datagram	64

- Mandatory bits included in IP packet headers according to Request for Comments (RFC) 791 (Postel, 1981a).
- Application data are appended to this IP control information, with the volume v of data being 1 <= v <= MTU, restricted by the Maximum Transmission Unit (MTU) of the link to which the node is attached.
- Additional field options may be appended to IP packets on a transmission-specific basis to supplement the information available.
- The overall minimum cost of IP packets is therefore the sum of those, and will be incurred by all packets on the end-to-end path passed between communicating nodes.
- IP modules implement the Internet Control Message Protocol (ICMP) defined in RFC 792 (Postel, 1981b), and ICMP messages are sent using the IP packet header (with an IP header Protocol field value of 1).
- ICMP reports problems with IP packet processing and can therefore send a range of error-reporting packets.
- This includes a Destination Unreachable Message (packet type 3), with fields included.
- Other ICMP packet types include Time Exceeded Message (type 11), Parameter Problem Message (type 12), Source Quench Message (type 4), Redirect Message (type 5), Echo (type 8) or Echo Reply Message (type 0), Timestamp (type 13) or Timestamp Reply Message (type 14) and Information Request (type 15) or Information Reply Message (type 16).

- The sizes of each of these packets are constant, with the code and type field varying as a function of the message type.
- Taking into account the packet header structure used by IPv4 and ICMPv4 protocols, there is a minimum of 144 bits in an IPv4 packet before application traffic is encapsulated and 96 bits in an ICMPv4 packet.
- Whilst IPv4 continues to be the most widely used version of the IP, IPv6 (Deering and Hinden, 1998) is used to support the rapid growth in the number of Internet users. In addition, IPv6 demonstrates a greater level of efficiency in its design, with the header fields restricted.
- IPv6 supports flexibility in its operation through use of an Options header (which contains the following fields: Option Type (8 bits), Option Data Length (8 bits) and Option Data (variable)).
- Option types include those which are read on a hop-by-hop manner and those read at the destination node only (both containing the fields Next Header (8 bits), Header Extension Length (8 bits) and Options (variable)).
- ICMPv6 (Conta, Deering, and Gupta, 2006) also demonstrates improved efficiency.
- Packets are divided into error and informational messages: The number of error messages is reduced from those provisioned in ICMPv4, and include only Destination Unreachable (type 1), Packet Too Big (type 2), Time Exceeded (type 3) and Parameter Problem (type 4) packets. Fields included within the Destination Unreachable packet, for example, include those used by ICMPv4.
- It is therefore through reduction in the number of packets used which improves its operational efficiency.
- With all header fields being mandatory, there is therefore a minimum of 320 bits in an IPv6 packet before application traffic is encapsulated and 96 bits in an ICMPv6 packet.

IPv6 packet header format



Field	Number of bits
Version	4
Traffic class	8
Flow label	20
Payload length	16
Next header	8
Hop limit	8
Source address	128
Destination address	128

Routing Information Protocol (RIP) Cost

- At the network layer, mandatory fields associated with packets transmitted using the Routing Information Protocol (RIP) according to RFC 2453 (Malkin, 1998) include Command (8 bits), Version (8 bits) and RIP Entry (between 1 and 25 entries).
- The RIP Entry is composed of the following fields: Address Family Identifier (16 bits), Route Tag (16 bits), IPv4 Address (32 bits), Subnet Mask (32 bits), Next Hop (32 bits) and Metric (32 bits).
- With each attribute being mandatory in all packets and the RIP Entry of variable length (with the potential of an array of information), there is a minimum of 176 bits of overhead in RIP packets.

AODV

- The routing protocol AODV (Perkins, Belding-Royer, and Das, 2003) supports a number of packet types, which include the Route Request (RREQ), Route Reply (RREP), Route Error and Route Reply Acknowledgement.
- With regard to the RREQ message as an example, packet fields include Type (8 bits), Flags including Join, Repair, Gratuitous RREP, Destination Only and Unknown Sequence Number (all 1 bit), Reserved (11 bits), Hop Count (8 bits), RREQ ID (32 bits), Destination IP Address (32 bits), Destination Sequence Number (32 bits), Originator IP Address (32 bits) and Originator Sequence Number (32 bits).
- With all attributes being mandatory in each packet, there is therefore 192 bits of overhead in the AODV RREQ message.

User Datagram Protocol (UDP) Cost

- At the transport layer of the stack, mandatory fields of User Datagram Protocol (UDP) packets according to RFC 768 (Postel, 1980) include the Source Address (16 bits), Destination Address (16 bits), Length (16 bits) and Checksum (16 bits).
- The volume of application data appended to each packet is controlled by the MTU of the link to which the node is attached.
- As each attribute is included in all packets transmitted using UDP, there is therefore 64 bits of overhead in UDP packets prior to the encapsulation of application data.

Transmission Control Protocol (TCP) Cost

- Mandatory bits associated with Transmission Control Protocol (TCP) packets according to RFC 793 (Postel, 1981c) include the Source Port (16 bits), Destination Port (16 bits), Sequence Number (32 bits), Acknowledgement Number (32 bits), Data Offset (4 bits), Reserved (6 bits), Control (6 bits), Window (16 bits), Checksum (16 bits), Urgent Pointer (16 bits) and the variably sized fields Options and Padding.
- As in the case of the IP and UDP protocols, the volume of application data appended to each packet is controlled by the MTU of the link to which the node is attached.
- With the header fields included in all packets transmitted using TCP, there is therefore a minimum of 160 bits in TCP packet headers prior to encapsulation of application data.

RTP

- The Real-Time Protocol (RTP) packet header according to RFC 3550 (Schulzrinne et al., 2003) contains the fields Version (2 bits), Padding (1 bit), Extension (1 bit), Contributing Source (CSRC) Count (4 bits), Marker (1 bit), Payload Type (7 bits), Sequence Number (16 bits), Timestamp (32 bits), Synchronization Source Identifier (32 bits) and CSRC Identifier (0–15 items, 32 bits each).
- With each attribute being used in all RTP packets, there is therefore a minimum of 96 bits in RTP packets prior to the encapsulation of application data.

Objectives of Green Network Protocols

Through exploring the range of header fields in a selection of commonly used protocols at different stack layers, potential opportunities to improve their efficiency have been identified.

1. Improving utilization of cross-layer detail between protocols unpacked at different stack layers;
2. Minimizing or eliminating redundancy in header detail;
3. Optimizing the protocols in their design through removal of support for older versions.

- A cross-layer approach promises the greatest improvements in energy efficiency (Almi’ani, Selvakennedy and Viglas, 2008), allowing problems and/or inefficiencies at each layer to be tackled in a consistent manner.
- When a protocol header is adapted, cross-layer compatibility will ensure that any attributes removed are not needed at other layers.
- Similarly, attributes incorporated for improved energy intelligence should be utilized at a maximum number of layers for optimum efficiency.

Meeting Objective 1

- Header detail may be better reused between stack layers in the case of the IP. The time to live (TTL) field is applied by the IP and also in the RSVP header, for example. As the network layer is traversed at each node, the attribute therefore need not be included in all headers but only in those used lower in the stack, and appended when being unpacked at the previous header layer.

Meeting Objective 2

- Inclusion of the ToS field in the IP header may be considered unnecessary at this layer. It describes the packet’s precedence, acceptable delay, volume of throughput and degree of reliability required. This detail may instead be gleaned from the TTL attribute which is also included in the header by default in an approach optimized for energy efficiency. On the other hand, the TTL can be captured from the ToS field and it need not be appended to the header instead. The nature of detail retained in these fields means that only one attribute is needed, not both.
- The need to include source and destination addresses in a range of packet headers used at different stack layers may also be questioned. Source and destination addresses are included, for example in MAC, IP, AODV, UDP, TCP and SNMP packet headers. Optimized UDP may, however, omit source and destination addresses from the packet header. This detail will be carried by the routing protocol and need therefore not also be replicated at the transport layer.
- The inclusion of a checksum within each protocol header can also be reconsidered for improved efficiency. Determined on an application-specific basis, it may be possible to optimize the inclusion of a checksum in all protocol headers, particularly in the case where the application can cope with a small degree of error, for the objective of optimizing communication energy efficiency.
- There may be redundancy in the header fields provisioned for IPv6. The hop limit field, for example, may be replaceable with the traffic class field only. The traffic class field can be used to indicate the acceptable delay associated with a packet stream (traffic classes remain undefined in RFC 2460), thereby removing the need to include both fields in the packet header.
- In the case of ICMPv6, there may be an opportunity to reduce the amount of redundancy associated with the protocol: Whilst there are fewer error message types used by this protocol in relation to those used by ICMPv4, it may be possible to restrict the range of error codes. With regard to the destination unreachable message, for example, there are seven optional error codes for reasons why the destination is unreachable. Three of these may, however, not be needed, including the options ‘Beyond Scope of Source Address’, ‘Address Unreachable’ and ‘Reject Route to Destination’, which could instead be replaced with the single error code, ‘No Route to Destination’.

Meeting Objective 3

- With regard to provision for support of updated versions of the protocol, this is the case with the Internet Group Management Protocol (IGMP) Version 3 defined in RFC 3376 (Cain et al., 2002) which also supports packet types associated with older versions of the protocol.

Green Network Protocols and Standards

Strategies to Reduce Carbon Emissions

- Business for Social Responsibility (BSR, 2009) suggests strategies to reduce carbon emissions at all stages of the business life cycle in general, from product manufacture to distribution.
- They suggest that carbon reductions are achievable by

  1. Enabling cleaner sourcing and manufacturing;

  2. Lowering emissions in transit;

  3. Enabling cleaner warehouse operations;

  4. Reducing transit distances;

  5. Removing nodes or legs;

  6. Reducing total volume and/or mass shipped;

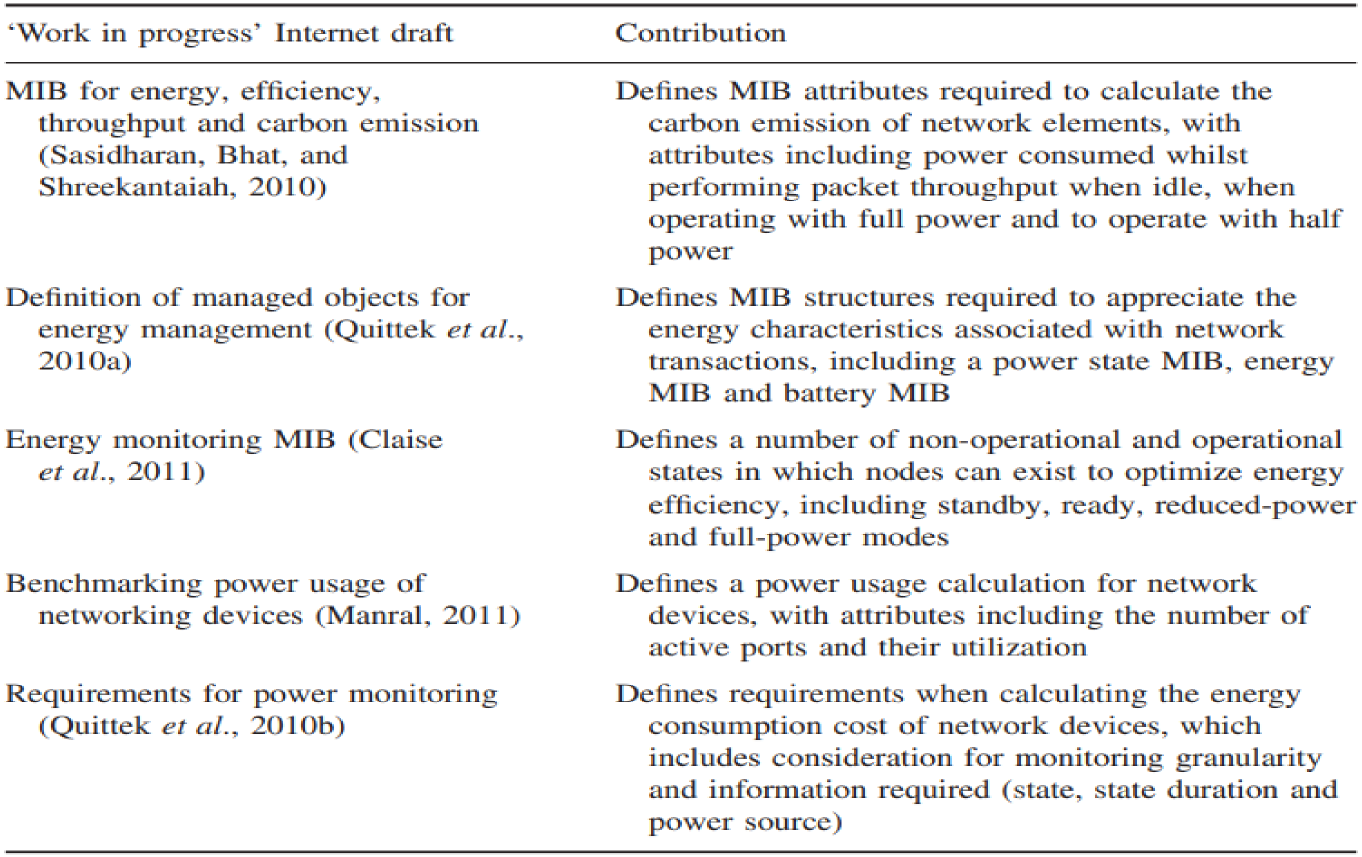
  7. Consolidating movements;

  8. Contributing to reductions elsewhere; and

  9. Increasing recycling and reuse.

- These techniques to reduce carbon emissions are not specific to telecommunications networks and consider carbon emitted during physical transportation of resources, development and production costs and onsite day-to-day operation.
- Whilst applied generically across businesses irrespective of their domain, we relate these to NGGN state-of-the-art strategies to demonstrate their versatility with regard to reducing carbon emissions in general, with processes involved during the communication of data having the same (albeit scaled-down) energy-associated impact.

Contributions from the EMAN Working Group



‘Work in progress’ Internet draft	Contribution
MIB for energy, efficiency, throughput and carbon emission (Sasidharan, Bhat, and Shreekantaiah, 2010)	Defines MIB attributes required to calculate the carbon emission of network elements, with attributes including power consumed whilst performing packet throughput when idle, when operating with full power and to operate with half power
Definition of managed objects for energy management (Quittek *et al.*, 2010a)	Defines MIB structures required to appreciate the energy characteristics associated with network transactions, including a power state MIB, energy MIB and battery MIB
Energy monitoring MIB (Claise *et al.*, 2011)	Defines a number of non-operational and operational states in which nodes can exist to optimize energy efficiency, including standby, ready, reduced-power and full-power modes
Benchmarking power usage of networking devices (Manral, 2011)	Defines a power usage calculation for network devices, with attributes including the number of active ports and their utilization
Requirements for power monitoring (Quittek *et al.*, 2010b)	Defines requirements when calculating the energy consumption cost of network devices, which includes consideration for monitoring granularity and information required (state, state duration and power source)

- The EMAN Working Group is involved in the conversion of ‘work in progress’ Internet drafts into formal RFC documents.
- In general, these drafts define management information base (MIB) structures designed to empower networks with energy awareness such that efficiency may be achieved.
- In MIB for Energy, Efficiency, Throughput and Carbon (Sasidharan, Bhat, and Shreekantaiah, 2010), for example, calculation of carbon emissions includes energy consumption, operational efficiency and utilization of each device attribute. Requirements for Power Monitoring (Quittek et al., 2010b) supplements Sasidharan, Bhat, and Shreekantaiah (2010) by defining requirements to perform energy calculations.
- This involves ensuring that all network components are monitored and that attributes collected include the current state and time spent in each state, total energy consumed at a device and since the last monitoring interval, and current battery charge, age, state and time when last used.
- The Energy Monitoring MIB (Claise et al., 2011) collects attribute details which include power cost per packet, duration of power demand intervals and maximum demand in a window.
- This Internet draft also considers compliance with MIB monitoring processes, with support for both reading and writing context from and to MIBs.
- Modes of improved operational efficiency are also suggested in this standard, and 12 power states may be applied to nodes in response to collected context.
- When related to the BSR’s principles, these strategies can be compared to enabling cleaner warehouse operations by improving understanding of the real-time environment and enforcing timely and appropriate actions to it.

Contributions from Standardization Bodies

- The European Telecommunications Standards Institute (ETSI) Environmental Engineering group defines techniques to monitor and control telecommunication infrastructure in response to collected context and predefined alarm conditions.
- Their drafts therefore define alarms, events and measurements necessary to provide the level of management required. In European Telecommunications Standards Institute (ETSI) (2010), AC Monitoring Diesel Back-Up System Control and Monitoring Information Model, for example the minimum range of events which should be monitored on a backup generator are defined, with alarms being raised if an undefined stop, start failure, fuel leakage or battery charger failure occurs.
- In ETSI, Environmental Engineering (2009), the monitored attributes of a DC power system control are defined.
- Alarms are raised when conditions include testing for battery failure, battery over-temperature and low-voltage output.
- These drafts further highlight the range of context which must be collected on an application-specific approach and the tailoring of alarms in relation to the domain in which management is applied.
- When compared to the BSR's strategies, integration of alarms such as those proposed by ETSI relate to lowering emissions in transit by suspending operations when environment conditions are insufficient to support it.

- IEEE 802.3, the Energy Efficient Ethernet (EEE) Study Group, is actively involved in reducing the power required to operate Ethernet technology.
- Primary contributions in the IEEE Standard 802.3az include a low-power state for activation during idle periods and times of low utilization (low-power idle (LPI)).
- This mode is applied in relation to link status and observed traffic flow.
- The standard also includes an alert signal which can be used to awaken those connections which have been sent to the sleep state when data arrive for transmission across an Ethernet link.
- When compared to the strategies proposed by the BSR, EEE relates to consolidating movements across primary links whilst suspending those across links which are not used as frequently.

Context Detail to Drive Energy Efficiency

- The EMAN working group has proposed MIB structures specific to the challenge of improved communication efficiency; ETSI defines alarms and measurements to control operation of power systems, and the IEEE defines strategies to optimize the power required to operate Ethernet technology.
- In addition, independent researchers propose solutions for application in individual domains and/or in response to a specific operational challenge at a specific stack layer, as in Ye, Heidemann and Estrin (2004) and Rhee et al. (2005).
- Taking into account developments in the field which provision information with regard to the nature of context required, the way in which it should be monitored, relevant evaluations and actions which may be applied, there is a research gap in that solutions have been provided in an ad hoc manner.
- In response to this, we have suggested that there are benefits to be achieved by applying domain-specific solutions to the collection and monitoring of context, evaluation and application of optimization strategies.

Network domain	Context used in each domain (per node)	Context used in each domain (in the wider environment between client and destination devices)
Data centre	*At an individual server within the data centre, context includes* server utilization, packet arrival rate (packets/second), power consumption rate (Watts/second), job completion rate (seconds), operational state (per node and per port), processing delay (seconds) and page faults (faults/page)	Bandwidth availability (bits per second), temperature (°C), power consumption rate (Watts per second) and operational state of neighbours (per node and per port)
Delay-tolerant network	*At an individual spacecraft deployed in deep space, context includes* tilt of solar panel (degrees), propagation distance from neighbours (seconds), critical activities, temperature (°C), line-of-sight connectivity with neighbours (true or false), residual battery capacity (units), received signal strength (dB) and operational state (per node and per port)	Wind speed (miles per hour), location of neighbours (x, y and z co-ordinates), residual battery capacity at neighbours (units), strength of signal arriving at neighbours (dB), operational state of neighbours (per node and per port), time of day, time of year, bit error rate (packets per second) and bandwidth availability (bits per second)
Mobile device	*For a mobile phone or laptop, context includes* backlight (% brightness), residual battery capacity (units), application type of service, device type, memory capacity (bits), device-critical activities, packet-sending rate (packets per second) and location (x, y and z co-ordinates)	Time of day (hours, minutes and seconds), bandwidth availability (bits per second) and location of neighbours (x, y and z co-ordinates)

Core	*At an individual router/switch in the network core, context includes* throughput (bits per second), utilization (%), operational state (per node and per port), energy cost per packet (Watts), packet-processing delay (seconds) and packet arrival rate (packets per second)	Bandwidth availability (bits per second), retransmission count at neighbours (packets per second), residual memory capacity at neighbours (bits) and bit error rate (packets per second)

Network domain	Context used in each domain (per node)	Context used in each domain (in the wider environment between client and destination devices)
Rural region	*At an individual networked device (client device or intermediary router), context includes* residual battery capacity (units), location (x, y and z co-ordinates), retransmission count (packets per second), packet transmission rate (packets per second), power cost per packet (Watts per packet) and packet arrival rate (packets per second)	Temperature (°C), bandwidth availability (bits per second), residual battery capacity at neighbours (units) and time of day (hours, minutes and seconds)
Smart home and office	*At an individual networked device, context includes* use of solar panel (true or false), device critical activities, operational state (per node and per port), time spent in state, energy cost per packet (Watts), time of last node sleep (hours, minutes and seconds) and sleep duration (seconds)	Bandwidth availability (bits per second), time of day (hours, minutes and seconds), location of nodes (x, y or z co-ordinates) and operational state of neighbours (per node and per port)
Wireless sensor network	*At an individual sensor, context includes* residual battery capacity (units), node location (x, y and z co-ordinates), operational state (per node and per port), propagation distance from neighbours (metres), temperature (°C), retransmission count (packets per second) and residual node memory (bits)	Temperature (°C), time of day (hours, seconds and minutes), location of neighbours (x, y and z co-ordinates), residual battery capacity at neighbours (units) and residual memory capacity at neighbours (bits)

- Context attributes are collected to drive intelligent decision making in terms of detail required on each individual node within the domain and also across the network within the wider environment.
- In exploring the problem domain in this way, optimization solutions in a range of environments which use different context and to which a range of contrasting evaluations should occur and actions can be applied are realized.
- Furthermore, in extension to standalone solutions identified in the literature, an integrated context-aware management solution which is cross-layer compatible across domains can also be developed (such as that proposed by the authors, the Energy-Efficient Context-Aware Broker or e-CAB; Peoples, Parr, and McClean, 2011), with a potential deployability and sustainability improvement in an approach similar to the TCP/IP and Open Systems Interconnection (OSI) protocol stacks upon which the Internet’s success to date has been built.