Journal of Electrical and Computer Engineering

Innovative Techniques for Power Consumption Saving in Telecommunication Networks

Guest Editors: Vincenzo Eramo, Xavier Hesselbach-Serra, and Yan Luo

Innovative Techniques for Power Consumption Saving in Telecommunication Networks

Journal of Electrical and Computer Engineering

Innovative Techniques for Power Consumption Saving in Telecommunication Networks

Guest Editors: Vincenzo Eramo, Xavier Hesselbach-Serra, and Yan Luo

Editorial Board

The editorial board of the journal is organized into sections that correspond to the subject areas covered by the journal.

Circuits and Systems

Communications

Signal Processing

Contents

Hindawi Publishing Corporation
Journal of Electrical and Computer Engineering
Volume 2014, Article ID 684987, 2 pages
http://dx.doi.org/10.1155/2014/684987

Editorial

Innovative Techniques for Power Consumption Saving in Telecommunication Networks

Vincenzo Eramo,[1] Xavier Hesselbach-Serra,[2] and Yan Luo[3]

[1] *Department of Information Engineering, Electronics, and Telecommunications (DIET), University of Roma Sapienza, 00184 Roma, Italy*
[2] *Department of Telematics Engineering, Technical University of Catalonia (UPC), 08034 Barcelona, Spain*
[3] *Department of Electronics and Computers Engineering (DECE), University of Massachusetts Lowell, Lowell, MA 01854, USA*

Correspondence should be addressed to Vincenzo Eramo; vincenzo.eramo@uniroma1.it

Received 2 March 2014; Accepted 2 March 2014; Published 5 May 2014

The increase in energy cost and the need to reduce the global greenhouse gas emission to protect our environment have stimulated the investigation of new innovative techniques for the energy consumption reduction in telecommunication networks.

Some studies show that ICT is today responsible for a fraction of the world energy consumption of about 4%, and this percentage is expected to double in the next decade. ICT is expected to play a major active role in the reduction of the worldwide energy requirements, through the optimization of energy consumption.

The purpose of this special issue is to study and evaluate the impact and potential exploitation of energy-aware innovative techniques for wired and wireless networks.

The special issue consists of seven papers. The first three papers focus on the definition and evaluation of techniques, based on device clock frequency reduction and turning off of links and nodes, for the power consumption saving in wired networks. The next three papers introduce and evaluate management schemes and technological solutions for reducing the power consumption in wireless networks. Finally, the last paper investigates the use of optical technologies in reducing the power consumption of switching nodes. Brief summaries of the accepted articles are listed below.

"*Design of a traffic-aware governor for green routers*" by A. Lombardo et al. focuses on routers that achieve energy saving by applying the frequency scaling approach. The authors propose an analytical model to support designers in choosing the main configuration parameters of the Router Governor in order to meet Quality of Service requirements while maximizing energy saving gain. A case study based on the open NetFPGA Reference Router is considered to show how the proposed model can be easily applied to a real case scenario.

"*Evaluation of power saving and feasibility study of migrations solutions in a virtual router network*" by V. Eramo et al. evaluates how the migration of virtual routers can lead to an energy saving. The mechanism consists in migrating virtual routers in fewer physical nodes when the traffic decreases allowing for a power consumption saving. After formulating the problem of minimizing the power consumption as a Mixed Integer Linear Programming, a heuristic is proposed to evaluate the power saving in real network and traffic scenarios. The authors also perform a feasibility study by means of an experimental testbed to evaluate the migration time of a routing plane based on QUAGGA routing software.

"*Facing the reality: validation of energy saving mechanisms on a testbed*" by E. Tego et al. focuses on the implementation of some techniques allowing for the turning off of router interfaces. Investigations on packet lost and delay are performed by means of an experimental testbed. The authors show that it is possible to dynamically adapt the network configuration to the changing load with no impact on packet loss and little increase in packet delay.

"*Energy-aware base stations: the effect of planning, management and femto layers*" by G. Koutitas et al. investigates

algorithms and techniques that can be applied on cellular networks to provide offered traffic proportional power consumption. Three different planning strategies and Base Station management schemes are used to investigate potential energy savings in the network. Furthermore, the paper shows how the introduction of a femtocell layer can improve energy saving, Quality of Service, and coverage providing more degrees of freedom to the mobile operator to adapt the power consumption in real time.

"*Smart power management and delay reduction for target tracking in wireless sensor networks*" by J. Feng et al. proposes a smart power management scheme in Wireless Sensor Network for target tracking application. Node sleeping strategies are introduced in the surveillance and tracking stages. Experimental results show that the proposed approaches are more power efficient with respect to traditional solutions and have a better capability of extending the network lifetime while maintaining short transmission delay in target tracking sensor networks.

"*A-LNT: a wireless sensor network platform for low-power real-time voice communications*" by Y. Fu focuses on the design of a lightweight low-speed and low-power wireless sensor platform for voice communications (A-LNT). The authors discuss the key elements for energy efficient node hardware design, low-power voice codec and processing, wireless network topology, and hybrid MAC protocol design. The efficiency in power consumption of A-LNT is studied with both simulation and analytical models.

"*Hybrid optical switching for data center networks*" by M. Fiorani et al. introduces a novel data center network based on hybrid optical switching (HOS). HOS combines optical circuit, burst, and packet switching on the same network. The proposed HOS network achieves high transmission efficiency and reduces energy consumption by using two parallel optical switches: a slow and low-power consuming switch for the transmission of circuits and long bursts and a fast switch for the transmission of packets and short bursts. By means of simulation and analytical investigations, the authors demonstrate that the proposed HOS data center network achieves high performance and flexibility while considerably reducing the energy consumption of current solutions.

Acknowledgments

The guest editors thank all the authors who submitted papers to the special issue, and they acknowledge all the reviewers for ensuring a high quality of the selected papers.

Vincenzo Eramo
Xavier Hesselbach-Serra
Yan Luo

Hindawi Publishing Corporation
Journal of Electrical and Computer Engineering
Volume 2014, Article ID 394376, 19 pages
http://dx.doi.org/10.1155/2014/394376

Research Article

A-LNT: A Wireless Sensor Network Platform for Low-Power Real-Time Voice Communications

Yong Fu,[1] Qiang Guo,[2] and Changying Chen[3]

[1] *Shandong Province Key Laboratory of Computer Network, Shandong Computer Science Center, Jinan 250014, China*
[2] *School of Management Science and Engineering, Shandong University of Finance and Economics, Jinan 250014, China*
[3] *Information Research Institute of Shandong Academy of Sciences, Jinan 250014, China*

Correspondence should be addressed to Yong Fu; yongfu0976@gmail.com

Received 1 November 2013; Revised 4 January 2014; Accepted 9 January 2014; Published 4 May 2014

Academic Editor: Vincenzo Eramo

Combining wireless sensor networks and voice communication for multidata hybrid wireless network suggests possible applications in numerous fields. However, voice communication and sensor data transmissions have significant differences, Meanwhile, high-speed massive real-time voice data processing poses challenges for hardware design, protocol design, and especially power management. In this paper, we present a wireless audio sensor network platform A-LNT and study and discuss key elements for systematic design and implementation: node hardware design, low-power voice codec and processing, wireless network topology, hybrid MAC protocol design based on superframe, radio channel allocation, and clock synchronization. Furthermore, we discuss energy management methods such as address filtering and efficient power management in detail. The experimental and simulation results show that A-LNT is a lightweight, low-power, low-speed, and high-performance wireless sensor network platform for multichannel real-time voice communications.

1. Introduction

WSNs (wireless sensor networks) consist of random or planned placed low-power wireless sensor nodes to monitor physical or environmental parameters [1]; these nodes are usually battery powered. WSNs have advantages of low power, low cost, self-networking, and no wiring or additional power supply needed. WSNs have been widely applied in environmental monitoring, intelligent transportation, automatic control, and so on [2]. Scientists and engineers are facing new demands and challenges with the deployments of WSN systems: precise real-time personal orientation, medical monitoring systems, access WAN and data fusion, and so forth. Voice communications have a wide range of potential applications in these WSN systems, such as staff regularly patrolling and examining, medical advice and counseling, broadcast and notification, and emergency voice communications. Although the research of WMSN (wireless multimedia sensor network) [3] has been carried out for ten years, high power and bandwidth requirements limit WMSN development [4]. In present, most WMSNs platforms are based on 802.11 platforms [5–7] and powered by high-capacity batteries or extern power supply.

In the last few years, multimedia components have become common and cheap with the rapid development of MEMS (microelectromechanical systems) and mobile Internet. Meanwhile, it becomes possible to achieve low-power and low bandwidth voice communication in WSNs with compression ratio improvement and power consumption reduction [8]. However, there are several problems that must be solved in combining voice communications with traditional WSNs: first, voice communication data and WSN data are significantly different in transmission features; audio data are real-time, which is different from ordinary WSN; voice communication needs much longer duration than sensor data transmission, meanwhile it occupoes the channel in communication. Moreover, audio sensor nodes are power-hungry and occupy massive bandwidth in data transmission. Furthermore, frequent high-speed real-time transmission of audio data poses challenges for WSN radio channel management, protocol design, hardware design, and energy management.

In this paper, we present a low-power wireless audio sensor network platform, which we called A-LNT: a lightweight low-speed and low-power wireless sensor network for voice communications, while considering WSN characteristics and hardware limitations. In Section 2, we will describe some existing WASN solutions. Then, in Section 3, we will discuss the hardware realization of A-LNT including node hardware components, power-management unit, and audio coding circuit. In Section 4, we will elaborate on the design of the MAC protocol including network topology, radio channel management, clock synchronization, and network management. In Section 5, we will describe our simulation and experiment procedures followed by reports of our results. At last, we evaluate and summarize A-LNT platform.

2. Related Works

There are a few suitable protocols and platforms for WASNs (wireless audio sensor networks) [9] at present [10–14]. Gabale et al. present a TDMA- (time division multiple access-) based MAC protocol LiT [10] and implement the MAC on an 802.15.4 platform Tmote; the evaluation of LiT shows quick flow setup, low packet delay, and essentiality for real-time applications; however, the speech coder chosen is G.723.1, and the coding bit is 6.3 kbps, but the speech codec power consumption is not considered; in 2012, a Lo^3 system based on LiT was reported [15], and it showed that such system bodes well with cost and power constraints in rural regions; the audio codec is SPEEX [16] with data rate of 5.9 Kbps, and the expected lifetime is 5 days. For most WSN applications, we need longer lifetime.

Li et al. study the audio element detection method in audio sensor network [11], and audio is treated as a special sensing data. In [12], speech codecs for high-quality voice over ZigBee applications are discussed. Zhao et al. design and implement an enhanced surveillance platform with low-power WASN; three kinds of audio sensors are discussed [13]. A voice network protocol based on session initiation protocol using TDMA/TDD MAC protocol using an IEEE802.15.4 PHY is present in [14], which is suitable for voice communications in both small- and large-scale networks. Further research of this group on full-duplex voice mixer for multiuser [17] and multiuser voice communications [18] is carried on.

Most of the above solutions are based on IEEE802.15.4/ZigBee protocol and an 8-bit RF SOC CC2430/CC2530; the ZigBee protocol is complex and huge, and the full protocol stack requires more than 100 Kbytes of flash and 5 Kbytes of ROM in CC2430/CC2530.

3. Hardware Realization

A-LNT is a WSN platform. It has the typical characteristics of WSNs: low power, self-organizing network, environmental parameter monitoring, and reliable data transmitting. Meanwhile, the platform could carry real-time voice communications without affecting sensing data transmissions. There are three types of nodes in our designed network: a central node (CNODE) for network establishing and management, sensor nodes (DNODEs) which are wireless terminals placed on target position or person for environmental monitoring and physiological parameters monitoring, and audio sensor nodes (ANODEs) which are wireless sensor terminals with audio communication functions. CNODE and DNODEs are typical nodes in WSNs, and ANODE is a new type of DNODE introduced by us for voice communications. All nodes are constructed by MCU, power management unit, RF transceiver, voltage monitoring unit, sensors, and batteries. ANODE and CNODE have additional parts for audio communications: audio processing unit, display unit, and user input unit. In order to simplify hardware design and embedded software programming, we choose the same MCU and RF transceiver for all nodes.

The MCU chosen is MSP430F2618, which is a 16-bit ultra-low-power RISC MCU; the MCLK is up to 16 MHz, and the wake-up time from low-power mode to active mode is less than 1us, which is suitable for dealing with frequent audio contents. There is an 8-channel 12-bit ADC (analog-to-digital converter) with internal reference, an internal temperature sensor, and 4 USCIs (serial communication interfaces) available in the chip, which could meet sensor interface needs for most WSNs. At present, we use the internal temperature sensor and one-channel ADC for voltage monitoring; there are 6-channel ADCs that are available for additional sensors. We chose CC2500 as the RF transceiver. It works at the ISM band of 2.4 GHz to 2.4835 GHz. The maximum wireless speed is 500 Kbps, and the current consumption is 17.0 mA at RX states, 21.1 mA@0 dBm at TX states, and 400 nA at sleep states.

DNODEs and ANODEs have different supply schemes. DNODEs are connected to batteries directly. It is an ideal way to power low-cost, low-power DNODEs as no energy loss is introduced by the power management unit. However, it is not suitable for powering ANODEs as audio codec and audio amplifier require low noise and a stable power supply. A high PSRR (power supply rejection ratio) LDO is necessary for ANODEs. Directly connecting LDO to batteries will increase current consumption and reduce available battery capacity. So a high performance step down DC-DC converter TPS62203 is added to the circuit in order to improve efficiency.

In practical application, users may want to turn off the terminal equipment when they finish their work. We use a low BISS (VCEsat) transistor PBSS5320T from NXP semiconductors and a small signal PNP transistor 9014 to realize a load switch. Although PMOS (P-Channel MosFet) transistors are popular in load switch designs, we chose a BISS transistor because it is ESD insensitive and has a constant VBE about 650 mV. So the voltage measurement circuit is easy to realize. The power control circuit workflow is as follows: when the batteries are connected to the board, the BISS transistor is off; when the tact switch $S2$ is pressed, the BISS transistor is ON, then the MCU turns on $Q3$, the board works normally, and the MCU monitors the voltage between $R4$ and $R5$. When $S2$ is pressed or the batteries voltage is lower than the threshold voltage for 30 seconds, the MCU turns off $Q2$

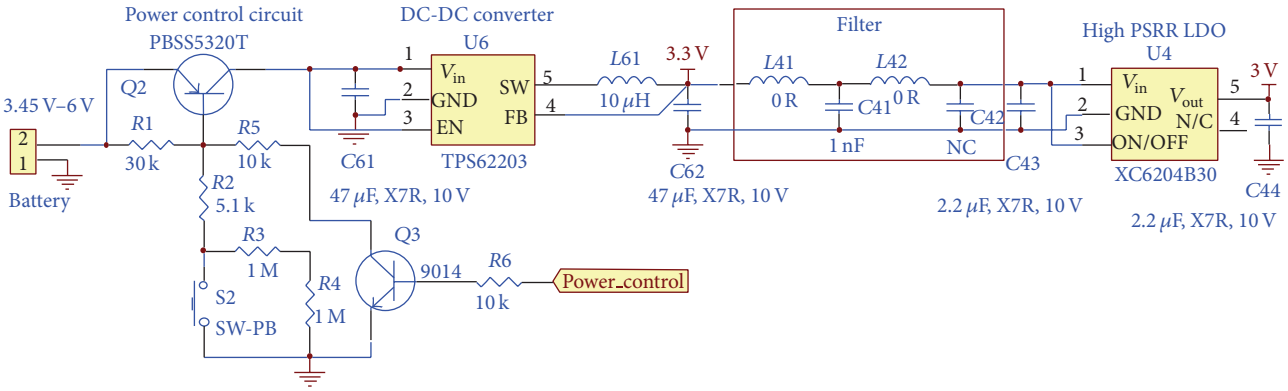


Figure 1: ANODE power management circuit.

and the board is powered down. Figure 1 shows the power management circuit schematic.

The last important part in hardware design is the audio processing unit. The audio processing unit consists of audio codec, microphone, audio amplifier, and communication interface. The audio codec algorithm should satisfy the following requirements: low power, low bit rate, and robustness for wireless communication. The CVSD (continuously variable slope delta modulation) algorithm meets all the above requirements and is an ideal solution for wireless voice communication; even the error bit ratio reaches 10%, and the MOS (mean opinion score) is greater than 3. CVSD algorithm is a simple algorithm based on PCM (pulse code modulation) algorithm, which has been widely applied in digital voice conferences and digital cordless telephones [19, 20]. The codec chip chosen is CMX649 [21], which is a low-power full-duplex codec; the typical operating current is 2.4 mA at 3.0 V, and the codec bit rate is 15.625 Kbps in the design. The codec transmits audio contents through SPI with MCU.

By now, we have finished the test-board hardware design. All test-boards in A-LNT are on a 2-layer FR-4 PCB where the board thickness is 1 mm. The DNODE test-board is a tiny board that consists of MSP430F2618, CC2500, 2 AAA batteries, and respective peripheral circuits. ANODE test-board is shown in Figure 2. The CNODE uses the same board as ANODE with different embedded software. We will discuss A-LNT MAC protocol design in the next section.

4. Protocol Design and Algorithm Realization

We divide the A-LNT protocol and software design into 3 parts: network topology, MAC protocol, and network management. Network topology is the foundation of the entire protocol design; it determines radio channel allocation and data transmission management strategies. MAC protocol is the most important part in A-LNT software design, which consists of hybrid channel access and management based on superframe, clock synchronization design, address filtering, address allocation rule, and packet priority setting. We will start from network structure design.

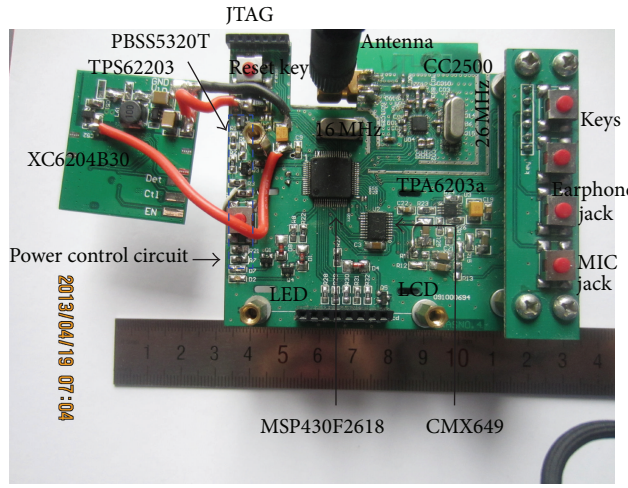


Figure 2: ANODE/CNODE test-board.

4.1. Network Structure Design. As we have mentioned above, A-LNT is based on WSN; it has typical characteristics of WSNs: low power, self-organizing network, environmental parameter monitoring, and reliable data transmitting. Meanwhile, the platform could carry out real-time voice communications without affecting sensing data transmissions. Sensing data requires reliable transmission, but latency is not critical. We ensure correct data transmission by applying the acknowledgment mechanism in WSNs. Voice communication is real-time, which needs strict clock synchronization. In addition, the real-time requirements and a large number of data transformations in voice communications determine that the network topology and protocol must be simple and efficient. Some packet loss and data error are acceptable in voice communications; acknowledgments are unnecessary and will degrade performance.

In a word, in order to meet the requirements of voice communications and sensing data transmission at the same time, the MAC protocol should be clock synchronous and ensuring two types of data noninterference. We designed the wireless network structure with considering the above factors; the network has a star topology as shown in Figure 3.

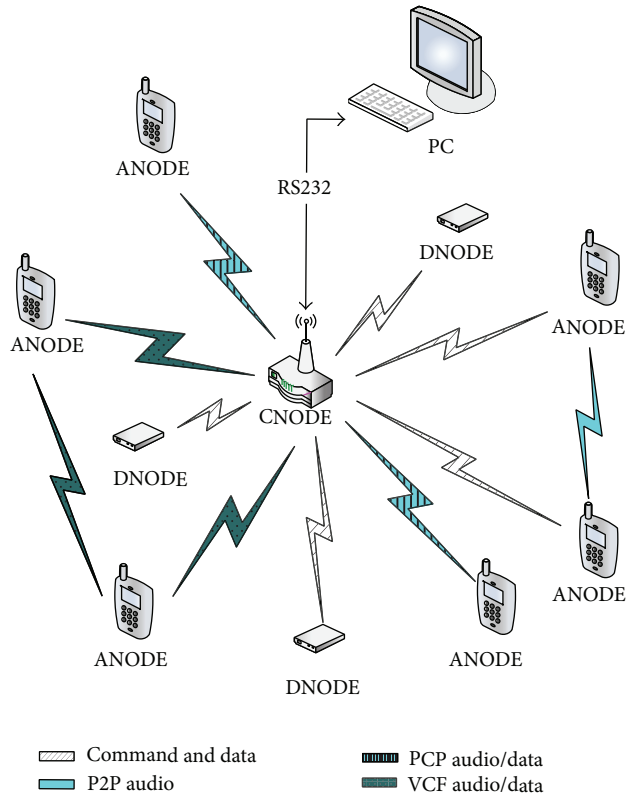


Figure 3: Network structure.

CNODE is in charge of network management, nodes management, and clock synchronizing. DNODEs measure environmental parameters and upload data to CNODE periodically. ANODEs upload sensing data in the same way of DNODEs. A-LNT supports three types of voice communications: in most conditions, voice communications between ANODEs should be peer-to-peer (P2P) in order to reduce wireless transmission pressure; if two ANODEs are too far to communicate directly, audio packets could be forwarded by CNODE (PCP); the last voice communication type is voice conference (VCF); in this mode, only one ANODE or CNODE is active while all other ANODEs are listening at one moment.

4.2. MAC Protocol Design. In order to realize the MAC protocol, we should determine clock synchronization frequency and the superframe time at first. In wireless multimedia networks, TDMA is an efficient and popular method to ensure QoS (quality of service) and maximize the use of wireless bandwidth [22, 23]. The transmitter sends multimedia contents at assigned time slot, while the receiver is listening, so they must be synchronous. Clock synchronization is critical in TDMA mechanism [24]. All nodes in A-LNT are synchronized when they join the network. However, the clock error will increase over time, which is caused by crystal tolerance and MCU clock accuracy. The clock error would lead to radio channel conflict, transmission failure, and system error. So periodic clock synchronization is necessary to maintain that the WASN works normally. The synchronization frequency should be a tradeoff decided by audio processing period and clock error. For a low-cost ±20 ppm crystal, in the worst case, the time error will reach 2 ms in 50 seconds; for a ±5 ppm crystal, the time will be 200 s, which is about 3 minutes.

The superframe time is decided by the audio sampling period and wireless period. In A-LNT, the codec bit rate is 15.625 Kbps; the audio codec generates bit stream continuous to MCU and demands the same number of bits from MCU in voice communications. In order to reduce complexity and power consumption, encoded audio content bytes generated by the codec should be less than the TX buffer size, which is 64 bytes. We design the superframe period T as 20.48 ms, which is also the audio sampling period; 40 bytes are sent to MCU in one T. In programming, we simplify operation by sending data to the codec; when receiving data from the codec, no additional timer or synchronization is required in this way. The audio data processing is mainly finished in the SPI interrupt function; in the interrupt function, MCU sends 1 byte decoded audio content when it receives 1 byte encoded audio from the audio codec.

ANODE sends encoded audio data periodically and receives wireless audio data from another ANODE in voice communications. In order to reduce hardware requirement and guarantee communication quality, we introduce four data buffers for cross access; two buffers are for storing received audio data, and the other two buffers are for storing encoded audio data. The maximum voice time delay is less than 2 times of T.

The superframe is divided into several time slots for sensing data transmission, network management packet transmission, and audio data transmission. The number of time slots is decided by the packet processing time. In order to get precise packet processing time, the packet processing time model is introduced as:

$$T_{\text{send}} = T_s + T_w + \frac{N_p + N_a}{B},$$
$$T_{\text{rev}} = T_r + T_m + \frac{N_p + N_a}{B}, \quad (1)$$

where T_{send} is the sending packet processing time; T_{rev} is the receiving packet processing time; T_s is the transmitter MCU processing time; T_w is the radio sending time; T_m is the receiver MCU processing time; T_r is the wireless receiving time; N_p is the packet payload length in bits; N_a is the preamble bits, sync word, and other data inserted automatically by CC2500; and B is radio transmission speed.

The packet processing time is decided by MCU main clock, data transmitting speed, packet length, and radio transmission speed. During voice communications, MCU main clock is 16 MHz, SPI speed is 4 Mbps, audio packet length is 46 bytes, and wireless speed is 500 kbps. T_{send} is 1.95 ms, which includes synthesizer calibration time 721 us. So the time slot for voice communication should be longer than 1.95 ms, and the time slot for network management and data transmission should be much longer than audio time slots for two reasons: at first, acknowledgment is usually necessary

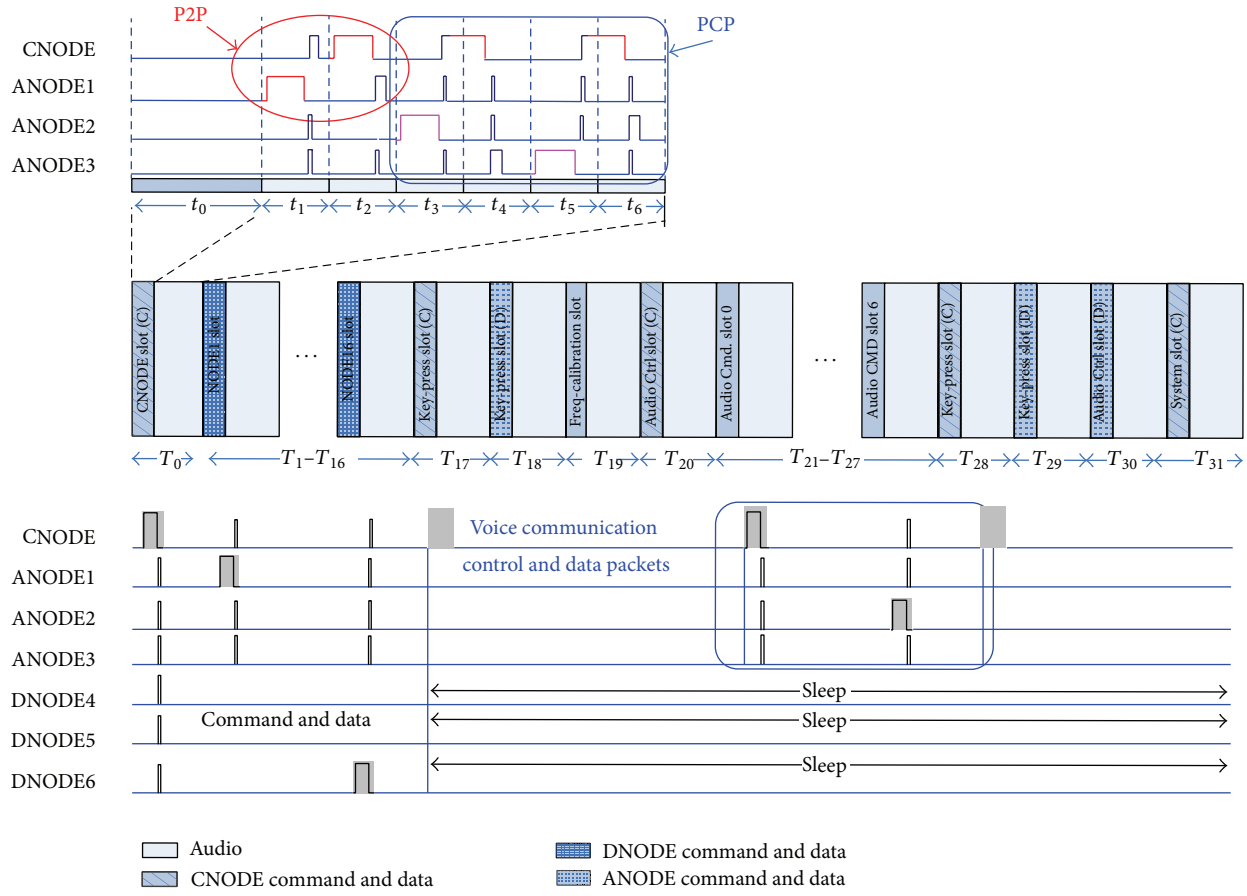


Figure 4: A-LNT channel allocation and address filtering diagram.

in network managements and it needs about twice the time; the other reason is that when the network works on low-speed mode, nodes do not have precise high-speed crystal or main clock; they need longer time for safety time interval and packet processing.

In multimedia communication applications, multimedia sensor nodes should be clock synchronized precisely with each other. However, high precision requires higher MCU clock, more expensive hardware, more current consumption, and shorter battery lifetime. In order to reduce power consumption, the audio processing units are shut down after voice communication finished and a superframe-based hybrid MAC protocol is introduced.

This MAC protocol mechanism consists of 4 key components: (1) the superframe is derived into data subframe and voice subframes. DNODEs listen to radio channels and send data only in data subframe; voice communications are carried out only in voice subframes. (2) The network adopts low time synchronization accuracy and lower node MCLK (main clock) to reduce energy consumption. The data subframe times are automaticly adjusted with network loads and CSMA/CA mechanism is adopted to manage radio channels. (3) When there are audio data transmissions, node MCLK is increased to work in full-speed mode and adopts high-precise time synchronization to ensure network performance. (4) Radio channels are allocated by center node using TDMA mechanism in voice subframes.

In detail, A-LNT consists of 1 CNODE, up to 16 ANODEs, and 64 DNODEs and an optional computer. The superframe T is 20.48 ms, and high main frequency is 16 MHz; the superframe is divided into 1 data subframe (t_0 = 6.08 ms) and 6 voice subframes (t_1–t_6 = 2.4 ms). When there are voice communications, ANODEs and CNODE listen to radio channels and send encoded audio content at specified voice subframes, and all DNODEs are in sleep. This platform supports up to 3-way P2P voice communications or 1-way PCP voice communication or a VCF including CNODE and all ANODES; the typical voice delay is less than 10 ms, and the time delay is less than 40 ms in the worst case. 32T compose a management cycle TT, the data subframe in the first superframe T_0 is a CNODE slot for network management. The data subframes in T_1–T_{16} are DNODE slots for network managements, network heart-beating, and sensing data transmissions. The data subframes in T_{17}–T_{31} are voice communication control and management slots. Every 6.55 s (i.e., 320T, the worst error at 20 ppm is 256 μs), CNODE broadcasts a polling packet POLL; all nodes should receive the packet, get time information, and adjust time to be in agreement with CNODE and then send reply packet ACK_POLL in specified slot, if there are sensor data need to

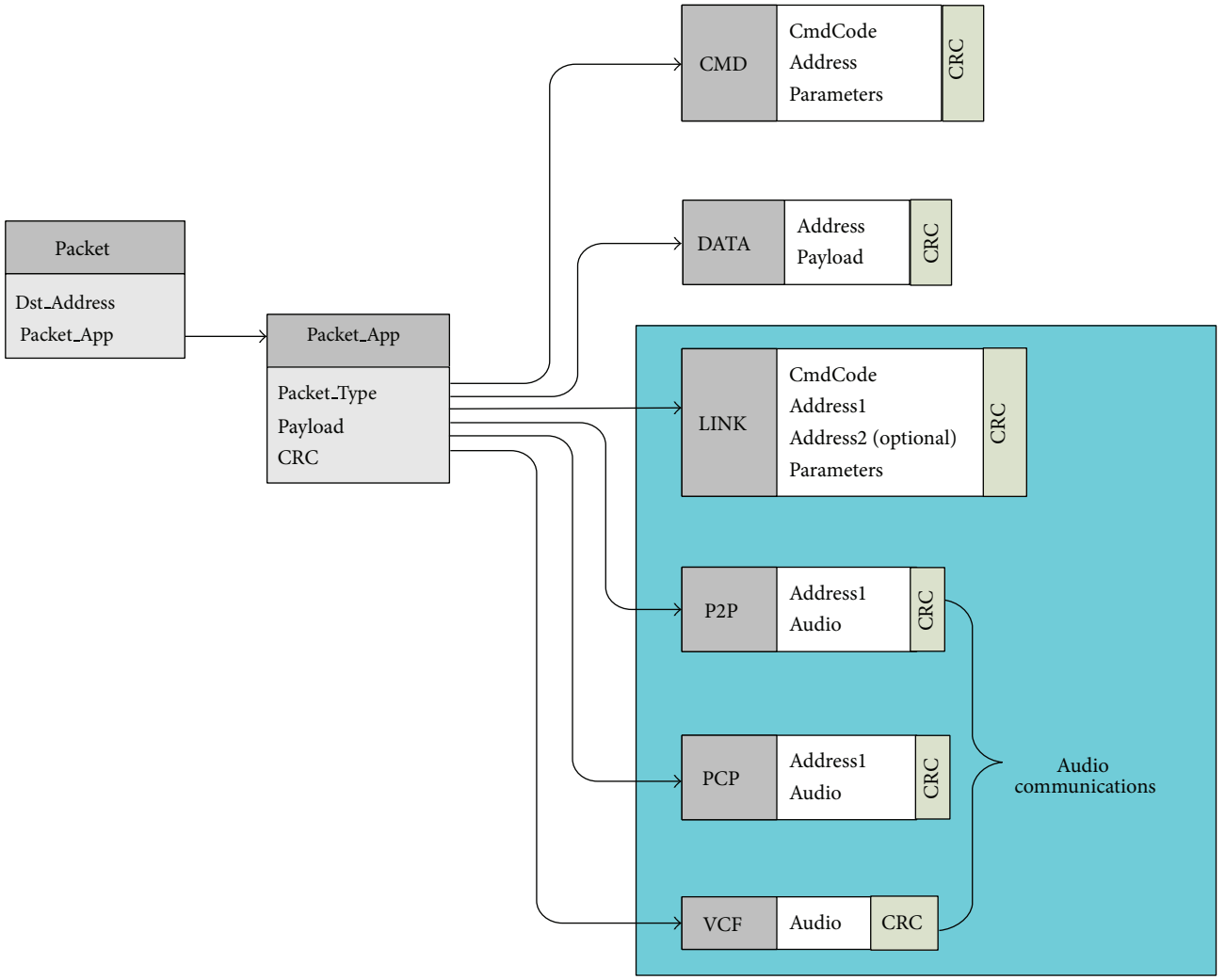


Figure 5: Packets structures.

be upload the nodes send them simultaneously by sending packet ACK_POLL_DATA instead of ACK_POLL; CNODE should send reply packet DATA_REVED to the node after receiving the packet immediately. For example, ANODE sends POLL at TT0. Nodes number 1–16 should send reply ACK_POLL/ACK_POLL_DATA in TT0, nodes number 17–32 should send reply ACK_POLL/ACK_POLL_DATA in TT1, and all replies should be finished in TT3. If CNODE did not receive ACK_POLL/ACK_POLL_DATA from one node for three successive management cycles, CNODE would delete the node. Priority design rules are as follows: CNODE has the highest priority, DNODEs have the highest priority in allocated slot, and other data are sent sequentially according to the priority within the data subframe. The high-speed crystal is shutdown when there is no voice communication, the node MCLK drops to about 2 MHz, and TT increases to 65.5 s (i.e., 3200 T; the worst error at 20 ppm is 2.56 ms). All nodes wake up in T_0 when the cycle is CNODE spooling cycle and go to sleep until it is time to send reply packet. When new node appears, it applies channel through CSMA/CA mechanism.

Address filtering is applied in A-LNT to reduce system total power consumption. In wireless network, all active nodes listen to radio channels; in most cases, only one node is the target node; other nodes receive useless packet and waste time to unpack and handle it. Address filtering is introduced to reduce wireless data processing time, which means that the wireless packet is unpacked and handled when address is matched; otherwise, the packet is abandoned. Address filtering can reduce processing time of that complete reception. This adaptive hybrid channel allocation method is an effective solution to the contradiction between multimedia communications and system power consumption. Figure 4 is the basic scheme of A-LNT channel allocation and address filtering protocol.

The other important pieces of information about A-LNT are as following.

The address allocation rule in the design is the following:

0X00: broadcasting address;

0x01: CNODE address;

0x02-0x0F: ANODE address;

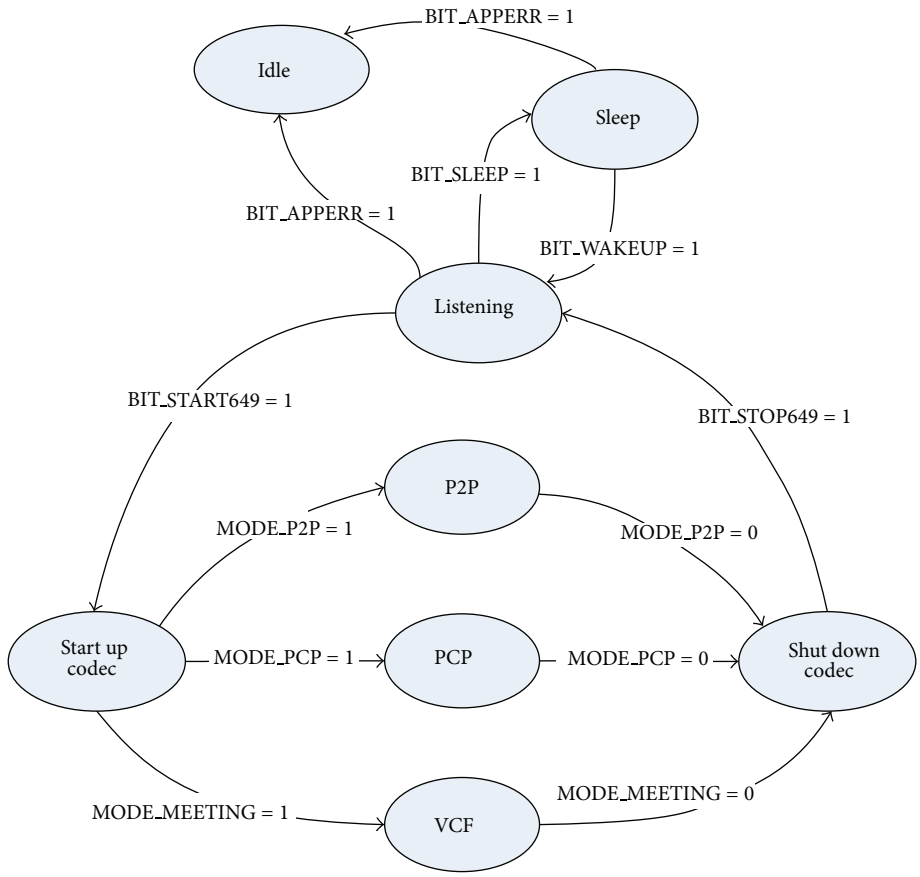


Figure 6: State machine of ANODE.

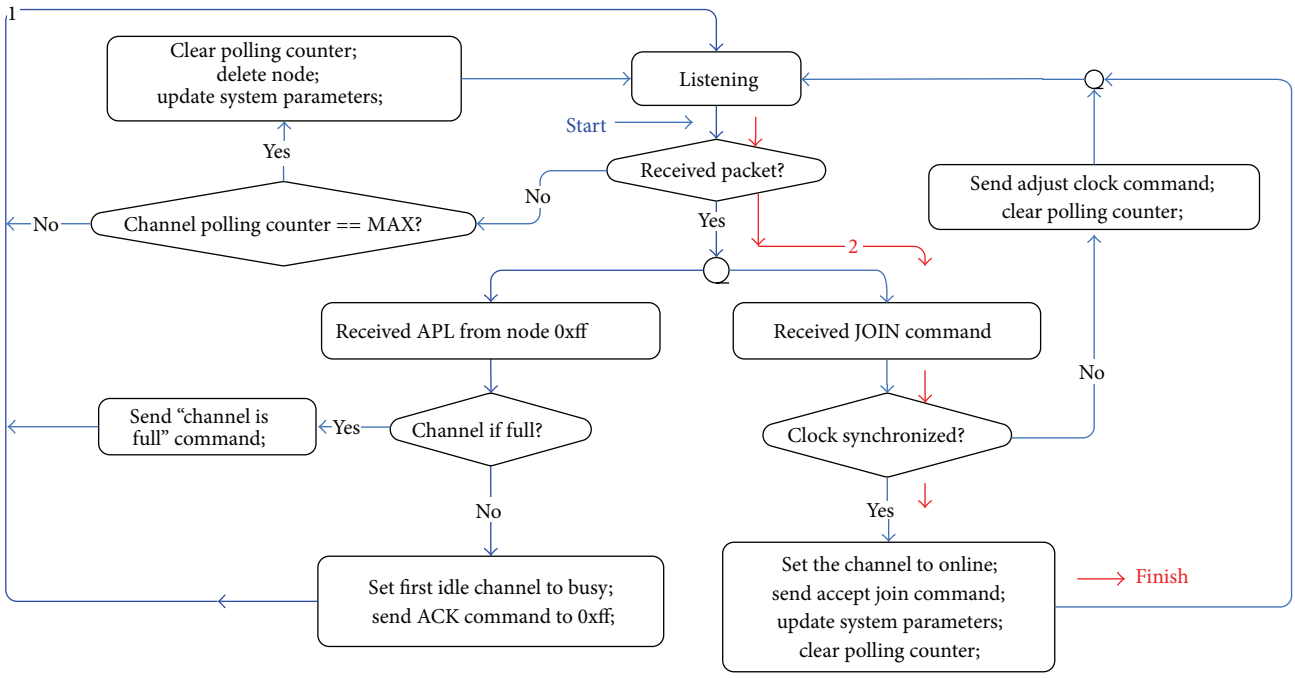


Figure 7: Simplified flowchart of node joining-CNODE part.

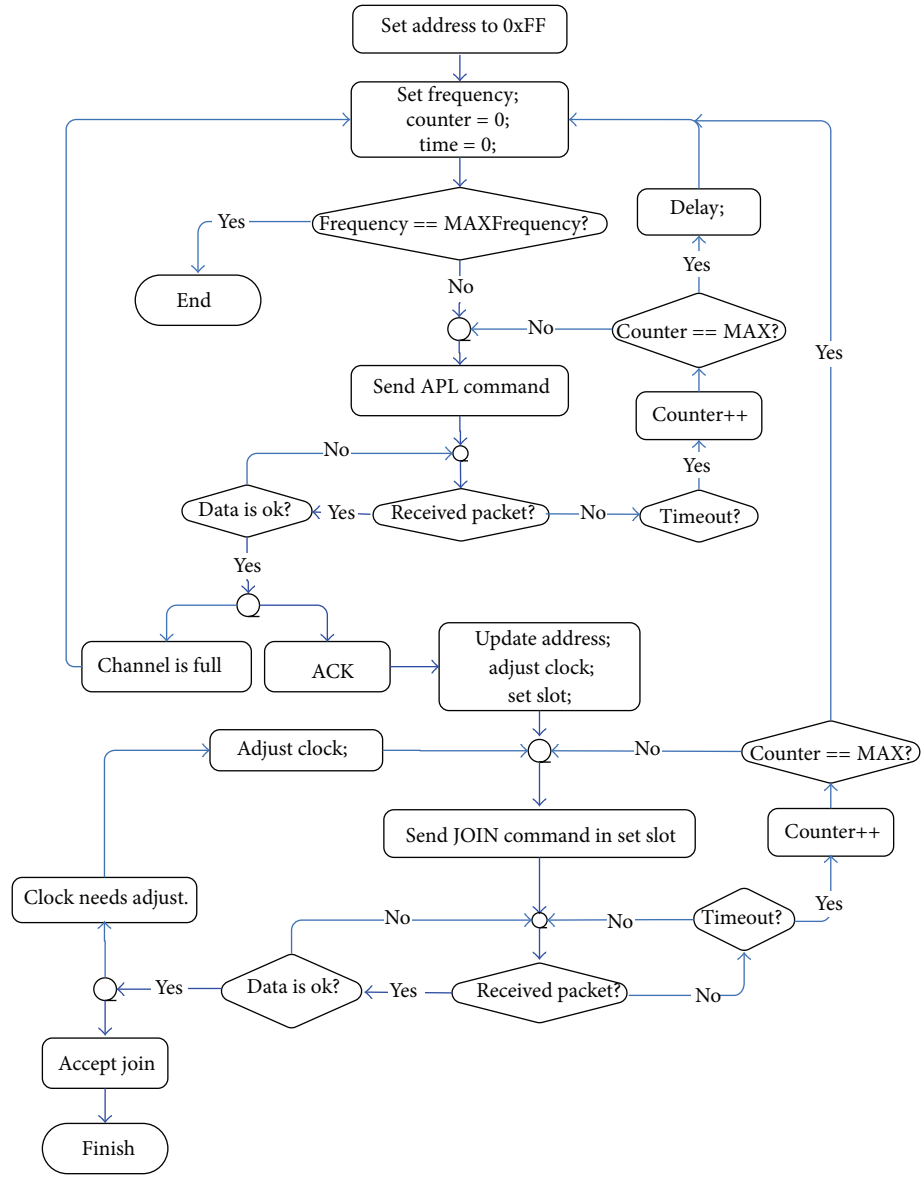


Figure 8: Simplified flowchart of node joining-ANODE/DNODE part.

0x10-0x7F: DNODE address;

0x80-0xEF: unused;

0xF0-0xFF: temporary address;

the wireless packets are divided into 6 types:

CMD, network management, priority 1;

DATA, sensing data and other data transmissions, priority 2;

LINK, voice communication link management and communication management, priority 3;

P2P, peer-to-peer audio data, priority 4;

PCP, peer-central-peer audio data, priority 5;

VCF, voice conference audio data, priority 6.

For more information about packet types, refer to Figure 5.

The state machine of ANODE is shown in Figure 6. ANODEs are sleeping in most times and go to listening mode at specified times. Listening mode could go to three voice communication modes, and audio codec only works in voice communication modes.

4.3. Network Management Protocol Realization. The last part of MAC protocol design is network management. The network management protocol takes on the role of the following:

(1) clock synchronization,

(2) radio channel management,

(3) nodes management.

```
(1)  Set RF channel = CHANNEL0
(2)   while (1)
(3)      Check RF channel
(4)      if RF channel is not available
(5)          If RF channel == MAX_RF_CHANNEL
(6)              Halt application.
(7)          else
(8)              RF channel +1;
(9)              continue;
(10)     else
(11)         break;
(12)     while (1)
(13)         Sleep;
(14)         Wake up for RF packet receiving or system event
(15)         if time == T_SENDING_SYNC // time for sending synchronization
(16)             Send POLL(ALL_NODES);
(17)         else if received RF packet
(18)              Unpack the packet, extract source address NODEA;
(19)              If packet type == ACK_POLL
(20)                   if the node is synchronized
(21)                        TOUT_NODE = 0; // clear timeout counter of the node
(22)              else if packet type == ACK_POLL_DATA
(23)                   if the node is synchronized
(24)                        TOUT_NODE = 0;
(25)                        Send DATA_REVED(NODEA);
(26)              else if packet type == APPLY
(27)                 if node-list is not full
(28)                        Assign a node number;
(29)                        Update node-list;
(30)                        Send ACK_APL(NODEA);
(31)                 else
(32)                        Send DENY(NODEA);
(33)              else if packet type == JOIN
(34)                   if node is synchronized
(35)                        Update node-list; // node has joined the network by now.
(36)                        Send ACP_JOIN(NODEA)
(37)                   else
(38)                        Send ACK_ JOIN (NODEA);
(39)              else if packet type == QueryPCP or QueryP2P // voice communication inquiry
(40)                  extract distinct address NODEB
(41)                  If free audio channel >0
(42)                      if (NODEB == CNODE)
(43)                          Send ACPPCP(NODEA);
(44)                          Start codec;
(45)                          Allocate audio channel;
(46)                          Start voice communication;
(47)                      else if. . . // more voice communication details, please browse the appendix.
(48)                          . . .;
(49)                  else
(50)                      Send DENY(NODEA);
(51)              else if packet type == Audio data
(52)                  . . .Packet processing; // please browse the appendix.
(53)              else if KEY pressed
(54)                      Response to KEY press;
(55)              else if time == T_CHECK_NODE // check nodes status
(56)                  for each node in node-list
(57)                      If TOUT_NODE > TOUTMAX
(58)                          Delete the node;
(59)              else
(60)                  . . .;
```

Algorithm 1

```
(1)  Set address = 0xFF
(2)  Set RF channel = CHANNEL0
(2)  while (1)
(3)       Set counter = 0;
(4)       Listen to radio channel for 100 ms
(5)       if RF packet received
(6)            Calculate system time
(7)       Set system parameter;
(8)       while (counter < CMAX)
(9)            Send APPLY(CNODE)
(10)           while (t < TTIMEOUT)
(11)               Listen to radio channel
(12)                   if ACK_APL received
(13)                       break;
(14)                   else if DENY received
(15)                           break;
(16)           if (t ≥ TTIMEOUT)
(17)               counter++;
(18)           else if DENY received
(19)               RF channel = unavailable;
(20)               break;
(21)           else if ACK_APL received
(22)               Jcounter = 0;
(23)               While (Jcounter < JMAX)
(24)                   Synchronize clock;
(25)                   Change address to new address
(26)                   Send JOIN(CNODE) at assigned time slot
(27)                   while (t < TTIMEOUT)
(28)                           Listen to radio channel
(29)                           if ACP_JOIN received
(30)                               break;
(31)                           else if ACK_JOIN received
(32)                               Synchronize clock;
(33)                               Send JOIN(CNODE) at assigned time slot
(34)                   if (network joined)
(35)                       Break;
(36)                   if (t ≥ TTIMEOUT)
(37)                       Jcounter++;
(38)       if (network joined)
(39)           Update system parameters;
(40)           break;
(41)       else if (RF channel is unavailable)
(42)           If RF channel == MAX_RF_CHANNEL
(43)               Halt application.
(44)           else
(45)               RF channel++;
(46)               continue;
(47)  while (1)
(48)       Sleep;
(49)       Wake up for RF packet receiving or system event
(50)       if time == T_ACK_POLL // time for sending ack-synchronization
(51)           if data needs to be sent
(52)               Send ACK_POLL_DATA(CNODE);
(53)           else
(54)               Send ACK_POLL (CNODE);
(55)       else if received RF packet
(56)           Unpack the packet, extract source address NODEA;
(57)           if packet type == POLL
(58)               Synchronize clock;
(59)               TOUT_NODE = 0; //clear timeout counter of the node
```

Algorithm 2: Continued.

```
(60)        else if packet type == DATA_REVED
(61)            Clear sent data information;
(62)        else if packet type == QueryP2P // voice communication inquiry
(63)            . . .; // more voice communication details please browse the appendix.
(64)        else if packet type == audio data
(65)            Packet processing; // please browse the appendix.
(66)        else if KEY pressed
(67)            Response to KEY press;
(68)        else if time == T_CHECK_NODE // CNODE status
(69)            If TOUT_NODE > TOUTMAX
(70)                reset application;
(71)        else. . .
(72)            . . .;
```

Algorithm 2

In our design, the work is mainly done by CNODE. It chooses an available radio channel and waits for radio packet receiving events. When a packet is received, CNODE does corresponding operation according to packet type. The network management process pseudocode is as shown in Algorithm 1.

The simplified flowchart of node joining is shown in Figure 7.

In order to join A-LNT, wireless node should seek an active CNODE, send APPLY to CNODE, get time information, and synchronize with CNODE. The process is as shown in Algorithm 2.

The simplified flowchart of node joining is shown in Figure 8.

By now, the A-LNT MAC protocol design is finished; it is simple and efficient, it consumes limited RAM and flash resources, and the details information is the following:

CNODE: 17 KB ROM, 0.5 KB RAM;

ANODE: 11 KB ROM, 0.4 KB RAM;

DNODE: 2.6 KB ROM, 0.1 KB RAM.

We would not discuss voice communication details here and have attached it to the end of the paper as it is tedious. In the next section, we will carry out experiments to verify A-LNT platform performance and discuss the results.

5. Results and Conclusion

At first, we measured the operating currents of three types of nodes (Table 1). The current consumption is mainly determined by audio unit and radio unit. It could be lower through reducing output power and receiving sensitivity of CC2500, turning the volume down of earphone, and powering down the LCM.

Then, we have studied the batteries lifetime in theory. In order to simplify calculation, we assume that the battery maintains OCV (constant open circuit voltage) 1.5 V and RI (the internal resistance) 150 mΩ. Battery capacity Q is 2300 mAh. The batteries are three alkaline batteries in series. The ANODE currents vary with operation mode: TX, RX, sleep, and audio. Average TX time tTX is 2 ms every 1000 T;

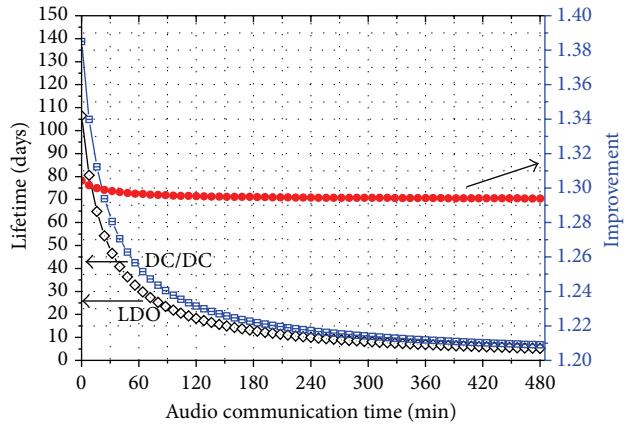


Figure 9: Battery lifetime versus voice communication time.

the RX time tRX is 3 ms every 1000 T. For boards with only LDO, the battery lifetime is given as follows:

$$T_{\text{Days}} = (Q \times 3600) \times \left(I\text{audio} \times t\text{audio} + I\text{rx} \times t\text{RX} + I\text{tx} \times t\text{TX} + I\text{sleep} \times t\text{sleep} + 3600 \times 24 \times I\text{ldo}\right)^{-1}, \quad (2)$$

where Ildo is supply current of LDO; in the design the LDO is XC6204B30 and Ildo is 70 uA.

For boards with DC/DC converters, battery lifetime is given as follows:

$$T_{\text{Days}} = (Q \times 3600) \times \left(I\text{audio}' \times t\text{audio} + I\text{rx}' \times t\text{RX} + I\text{tx}' \times t\text{TX} + I\text{sleep}' \times t\text{sleep} + 3600 \times 24 \times I\text{ldo}'\right)^{-1}, \quad (3)$$

where

$$\text{Ii}' = \frac{\text{Ii} \times V_{\text{out}}}{(\text{OCV} - \text{RI} \times \text{Ii}') \times \eta} \quad (4)$$

and $V_{\text{out}} = 3.3$ V, $\eta = 95\%$.

Table 1: Current summary.

Node type	TX (mA)	RX (mA)	Audio (mA)	Average current (mA)	Sleep mode (mA)
CNODE	21.4	19.6	27	53	3.2
ANODE	21.4	19.6	27	52	0.9
DNODE	21.4	19.6	—	22	0.6

Table 2: Processing times of different packet lengths and hardware.

Time (us)	Address match	MCLK (MHz)	SPI speed (Kbps)	Payload length (bytes)
200	Yes	16	4000	5
45	No	16	4000	5
900	Yes	2	500	5
160	No	2	500	5
350	Yes	16	4000	12
45	No	16	4000	12
1500	Yes	2	500	12
160	No	2	500	12
1050	Yes	16	4000	46
45	No	16	4000	46

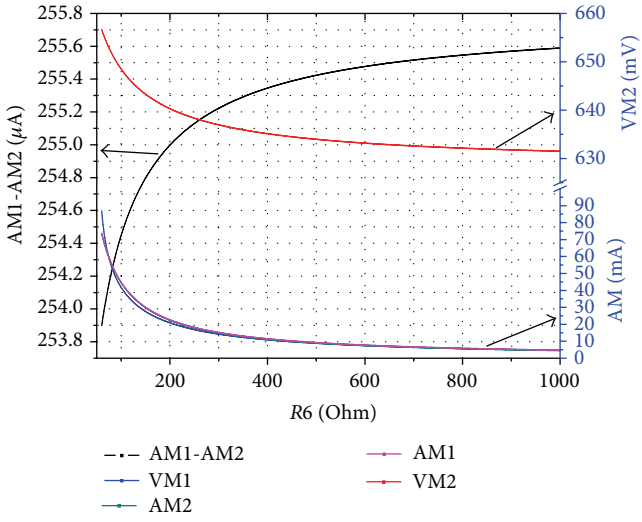


Figure 10: Voltages and currents versus loads.

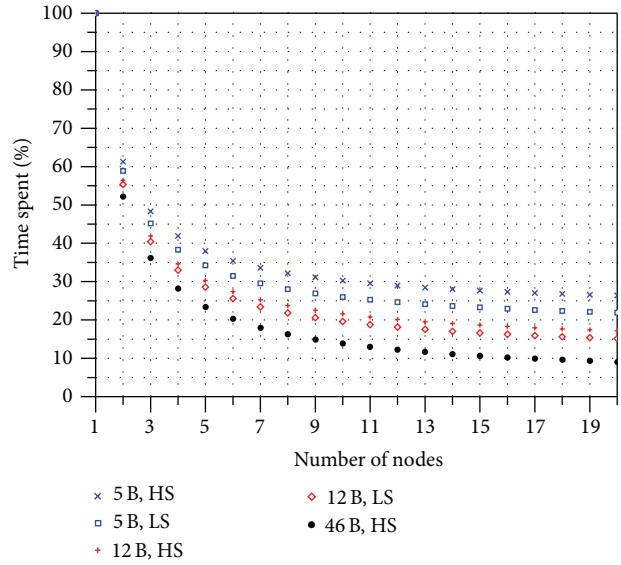


Figure 11: Time spent versus number of nodes.

Figure 9 shows the calculation results. The DC/DC converter extends battery lifetime by greater than 29%. If voice communication time is 30 minutes per day, ANODEs could work for more than 60 days without changing batteries. It is also possible to serialize more batteries or use high voltage batteries extending node working time.

The minimum input voltage is calculated by the following equations:

$$V_{\text{in min}} = V_{\text{out max}} + \text{IL}_{\text{max}} \times \left(r_{\text{ds}}\,(\text{ON})\,\text{max} + \text{RL}\right),$$
$$\text{IL}_{\text{max}} = I_{\text{out max}} + \frac{V_{\text{out}} \times \left(1 - \left(V_{\text{out}}/V_{\text{in}}\right)\right)/\left(L \times f\right)}{2}, \quad (5)$$

where $I_{\text{out max}} = 53\,\text{mA}$; $V_{\text{out}} = 3.3\,\text{V}$; $L = 10\,\mu\text{H}$; $f = 1\,\text{MHz}$; $r_{\text{ds}}(\text{ON})\,\text{max} = 670\,\text{m}\Omega$; the power inductance is CDRH5D28NP-100N from Sumida, and $\text{RL} = 65\,\text{m}\Omega$. The result shows that $V_{\text{in min}}$ is less than 3.4 V, for three alkaline batteries in series; the node could extract almost all energy.

The load switch circuit simulation is carried out by TINA-Ti 9.0; the results are shown in Figure 10. The current consumption of BISS transistor is about 255 uA, and the VBEsat is about 50 mV when the load current is 50 mA. Node shutdown current consumption is only 2.21 uA. The power management circuitry has virtually no impact on the node power consumption.

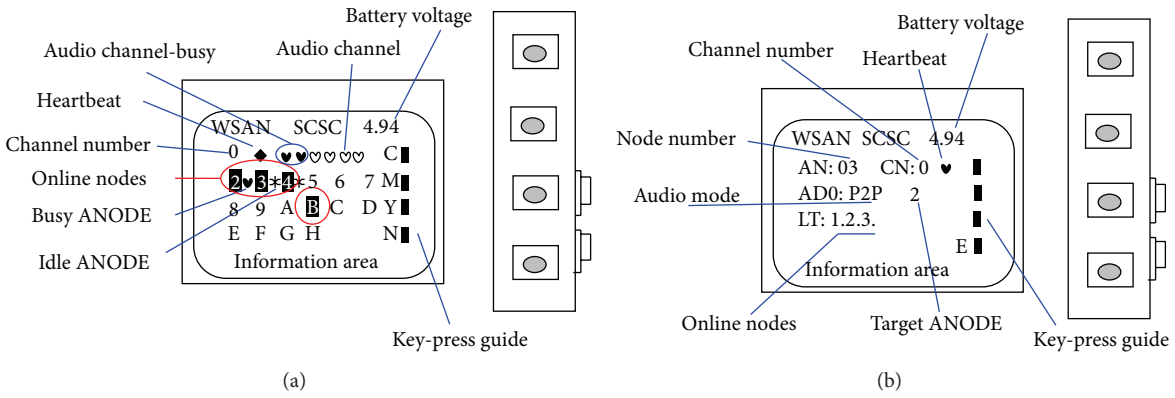


Figure 12: LCD screen of CNODE (a) and ANODE (b).

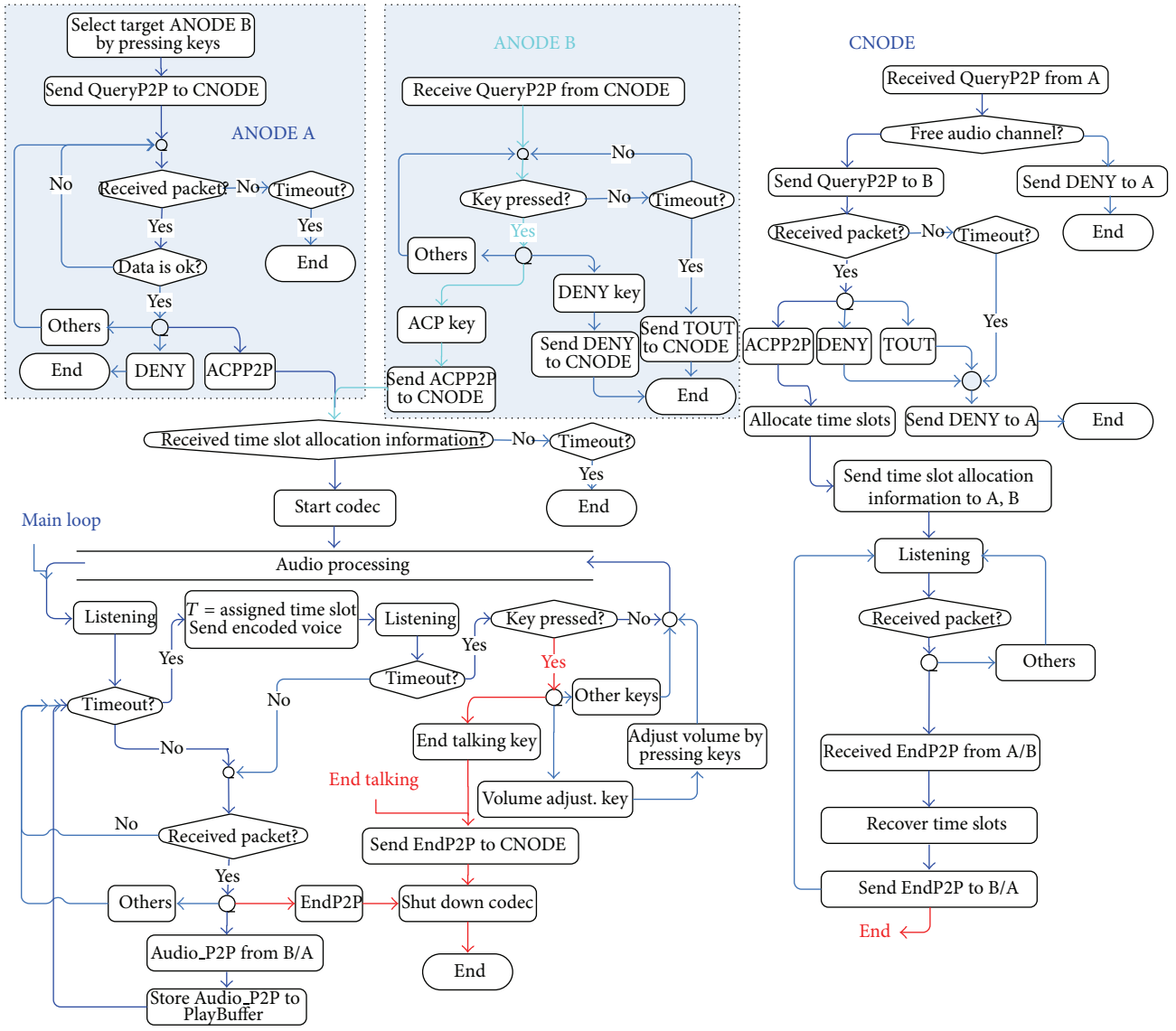


Figure 13: P2P communication flowchart.

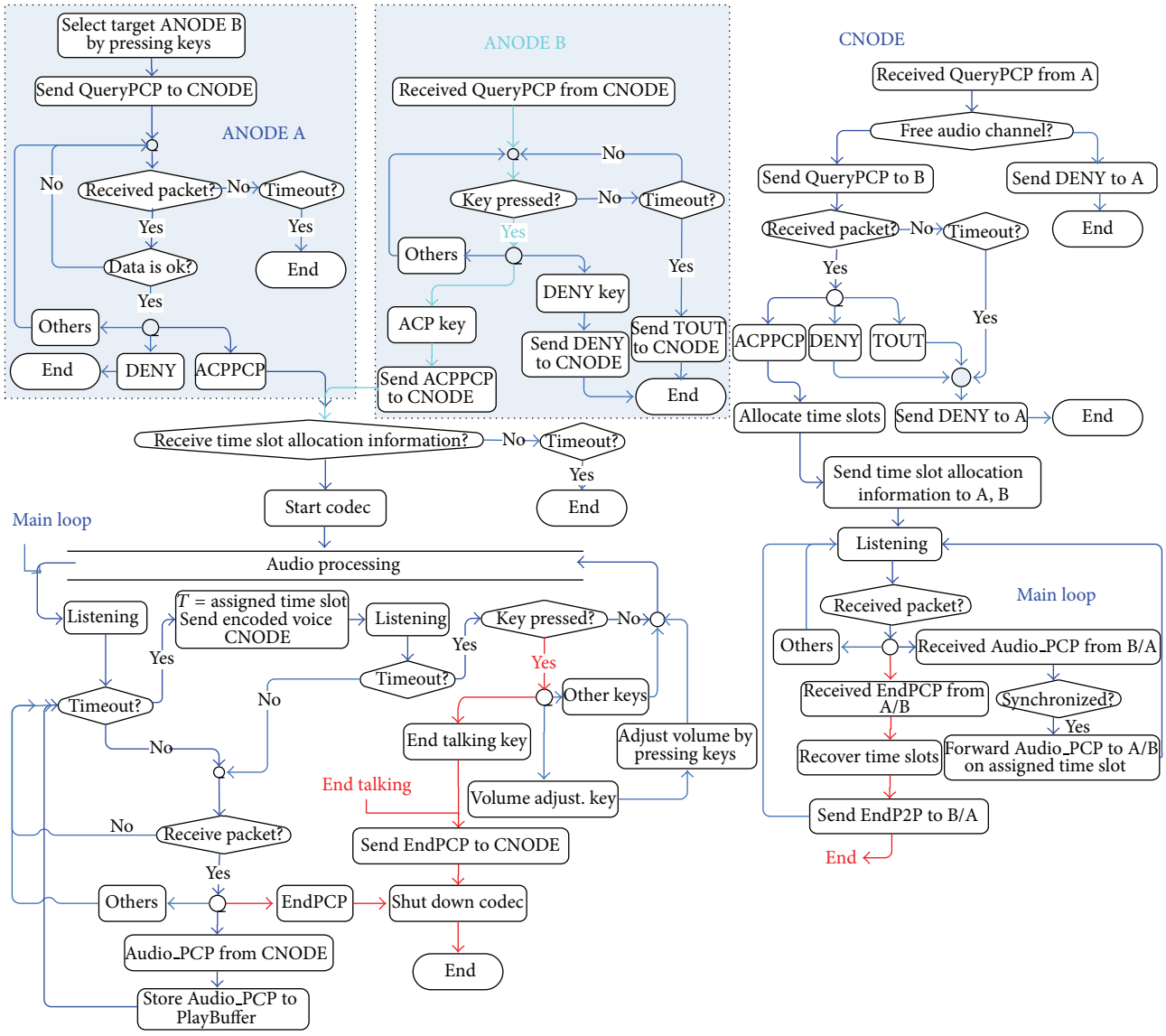


Figure 14: PCP communication flowchart.

The communication distance between ANODEs is longer than 70 meters indoors and 120 meters outdoors. The results are measured under the following conditions: line of sight, being about 1.5 meters above ground, and the nodes being placed on a table or carried by person.

At last, the address filtering performance is studied by measuring RF packet processing times. We measure packet processing times with a pair of nodes. Node A sends the same packets every 100 ms for 200 times. At the same time, node B stays in RX states when the RF packet is detected; the timer starts counting until the packet is received and processed, and then the timer count is stored into an array. At last, the processing time is calculated by averaging all 200 counts. The processing times of different packet lengths are measured as shown in Table 2.

The timing accuracy is 5 us in the above measures. The total processing time saving with parameters in Table 2 versus number of nodes is shown in Figure 11.

Where B means bytes, HS means "high speed" and LS means "low speed." It can be seen from Figure 11 that the processing time of all active nodes in WSN reduces with the node number increasing and packet length increasing. When the number of ANODEs reaches 6, the time spent is reduced to 20% of that without address filtering. It is an efficient method to reduce network power consumption.

In conclusion, we have presented a low-power WASN platform A-LNT from hardware realization to protocol design. The network has a star topology and three voice communication modes: P2P, PCP, and VCF. The audio codec is CVSD 15.625 kbps; cross accessing and data buffer pool

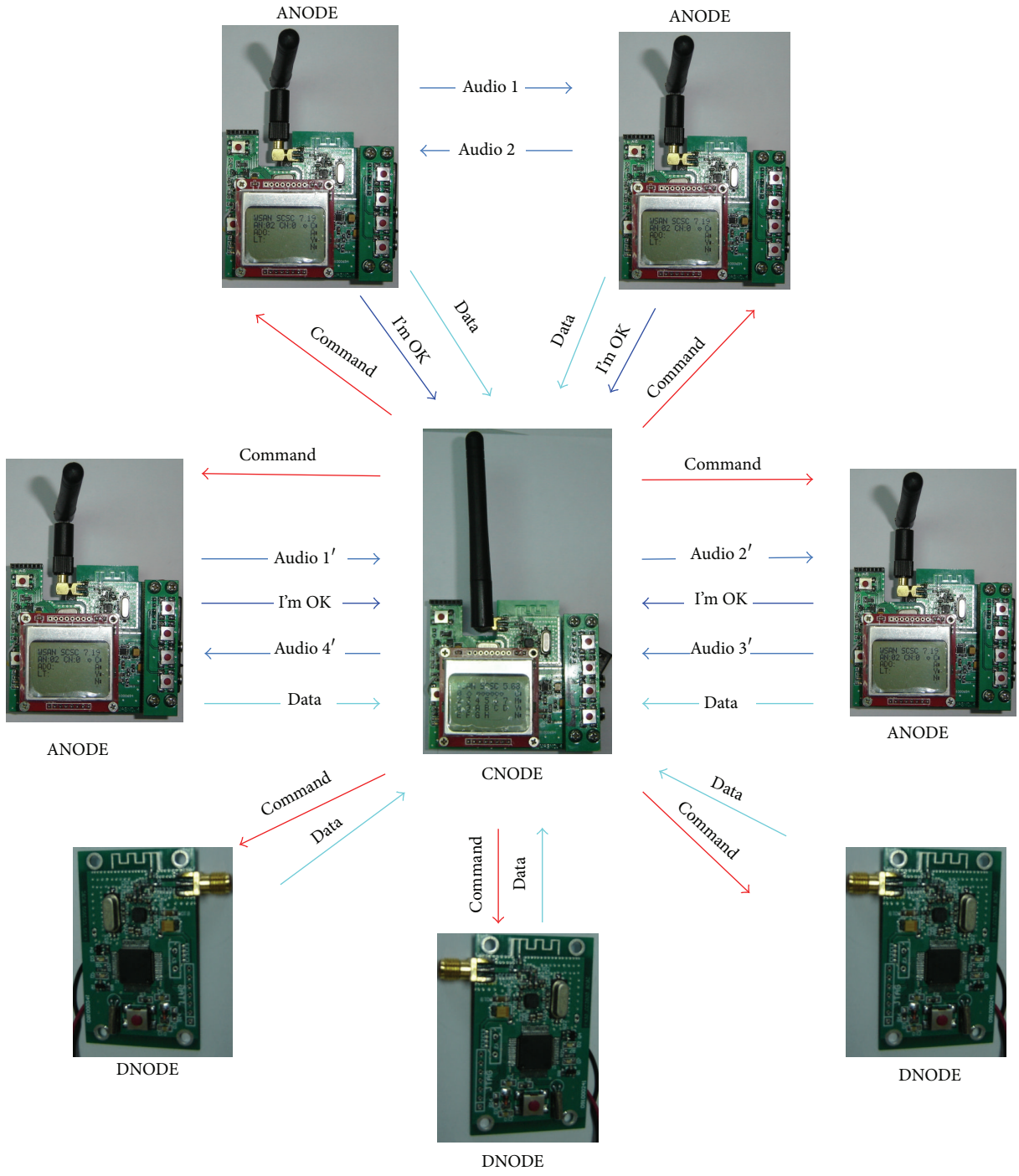


Figure 15: Schematic diagram of P2P and PCP.

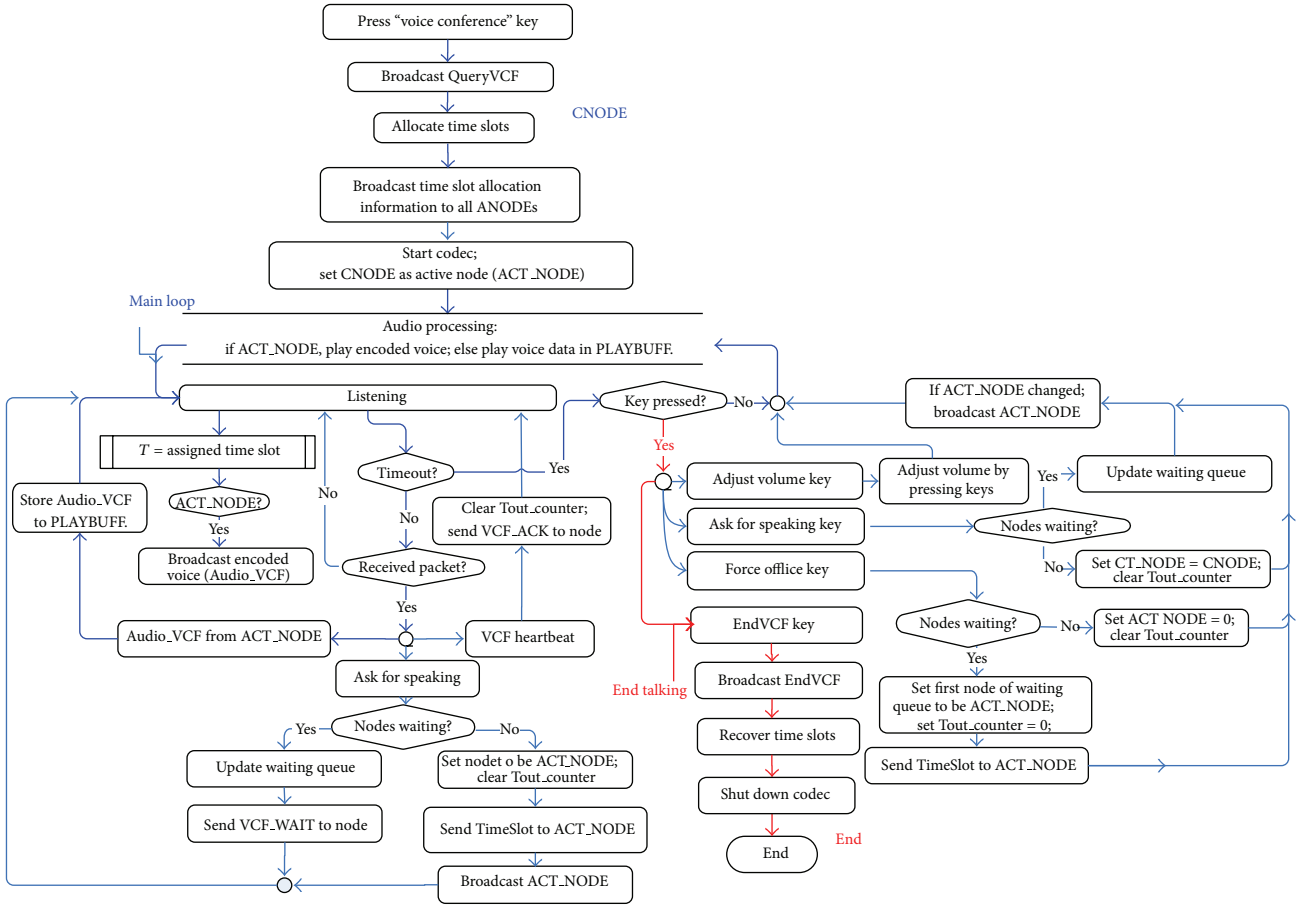


Figure 16: VCF communication flowchart—CNODE.

are introduced to reduce hardware requirement and power consumption. The efficient power management method for A-LNT is studied by simulation and mathematical theory in detail. We also designed a hybrid MAC protocol based on superframe. The superframe setting is studied by considering the audio sampling period, wireless packet processing, clock error, and voice communication channels. The superframe time is 20.48 ms; it is divided into a time slot (t_0, 6.08 ms) for network management and data transmission and 6 audio time slots (t_1–t_6, 2.4 ms). Address filtering and sleeping in specified time slots are applied to reduce wireless packet processing time and power consumption. MAC protocol key elements such as clock synchronization and network management are discussed also. The result shows that A-LNT is a low-power, low-speed, and high-performance WSN platform. It consists of up to 16 ANODEs, 64 DNODEs, and 1 CNODE. The audio channel capacity is 3 real-time two-way voice communications or audio conference including all audio nodes at the same time. And the voice delay is less than 40 ms. It suggests possible applications to emergency voice communication, audio/sound sensor network, health monitoring system, and so forth.

Our future work is concerned with increasing voice communication channel number and reducing power consumption. In detail, we are looking for new audio codec with low bit rate and low power, new type of wireless transceiver with fast switch speed from idle/sleep status to TX/RX status, high-speed and low-power wireless transceiver, and wireless protocol design.

Appendix

A. Voice Communication Process

There are three voice communication modes: P2P, PCP, and VCF. All communication managements are operated by CNODE. The voice communications are initiated and managed by pressing keys according to LCD screen guidance information (Figure 12). We will discuss these types of voice communications in detail.

A.1. P2P Mode. ANODE A selects target ANODE B by key pressing and sends "QueryP2P" command to CNODE; CNODE checks whether there are free audio time slots and whether ANODE B is free. If both conditions are met, CNODE forwards "QueryP2P" to ANODE B and waits for answering. If ANODE B denies the query or timeout, CNODE sends 'DENY' command to ANODE A and the

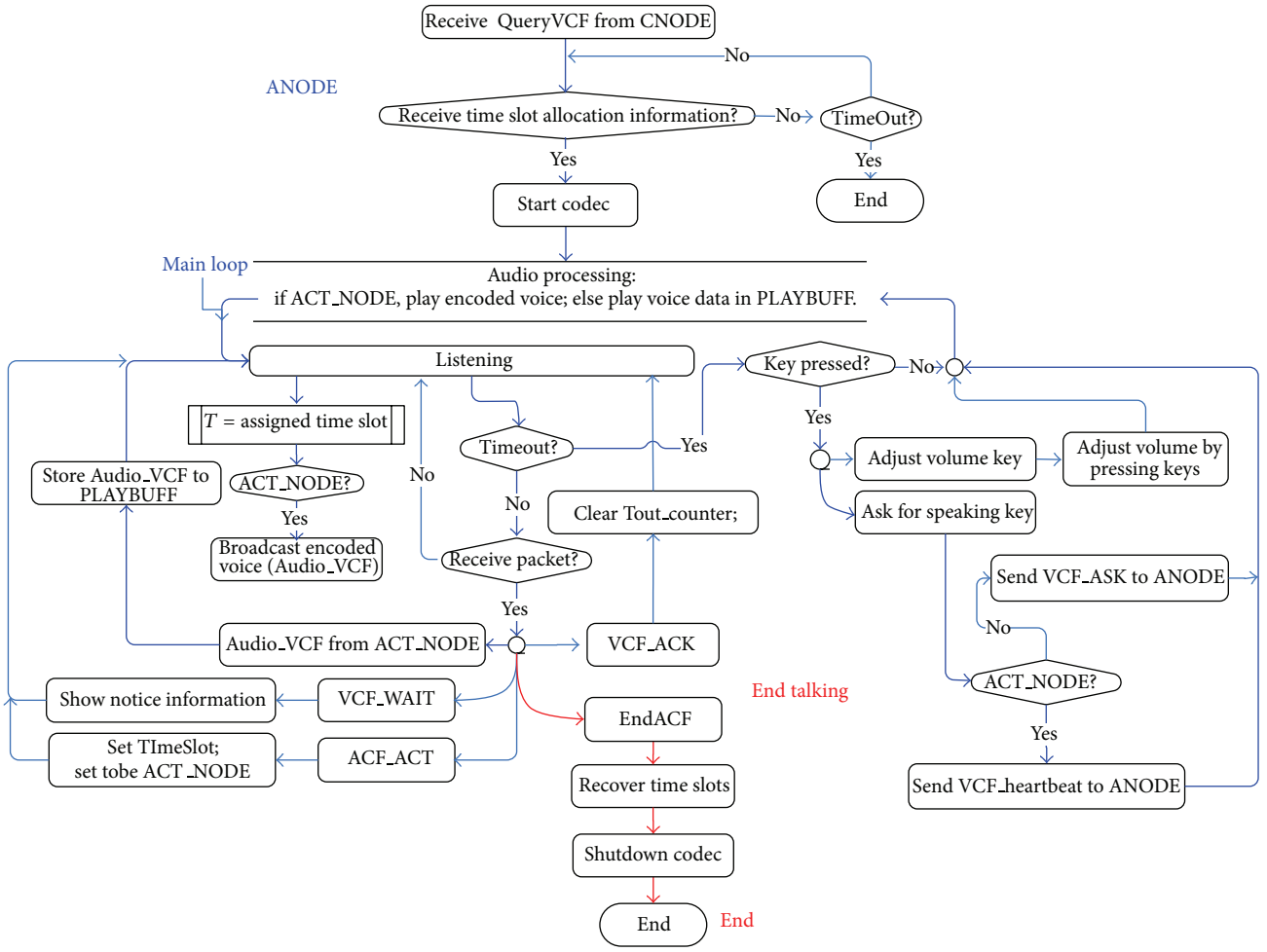


Figure 17: VCF communication flowchart—ANODE.

voice communication halts. If ANODE B accepts the query, CNODE allocates time slots to ANODEs A and B and checks communication status during voice communication. ANODEs A and B receive time slot information and then they start audio codec and send audio contents on assigned time slots. In P2P communication, ANODEs A and B send heartbeat packets to CNODE periodically. ANODE A/B could end communication by pressing "End communication" key. Meanwhile, the node sends "EndP2P" command to CNODE; CNODE withdraws the time slots and forwards the command to the other ANODE. The P2P communication is ended. The simplified flowchart is shown in Figure 13.

A.2. PCP Mode. PCP mode is similar to P2P mode; the differences are that CNODE allocates 4 time slots and all audio contents are forwarded by CNODE. The simplified flowchart is shown in Figure 14.

The schematic diagram of P2P and PCP is shown in Figure 15.

A.3. VCF Mode. VCF is initiated by CNODE. Any ANODE that wants to speak should send VCF_ASK command to CNODE; CNODE checks if there is an active speaker and sends respect reply to the asker. The VCF could be ended only by CNODE. The simplified flowchart is shown in Figures 16 and 17.

The schematic diagram of P2P and PCP is shown in Figure 18.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This paper is supported by the Shandong Provincial Foundation for Outstanding Young Scientists (Grants nos. BS2012DX035 and BS2011DX031). The authors also thank the anonymous reviewers and the editor for their valuable comments.

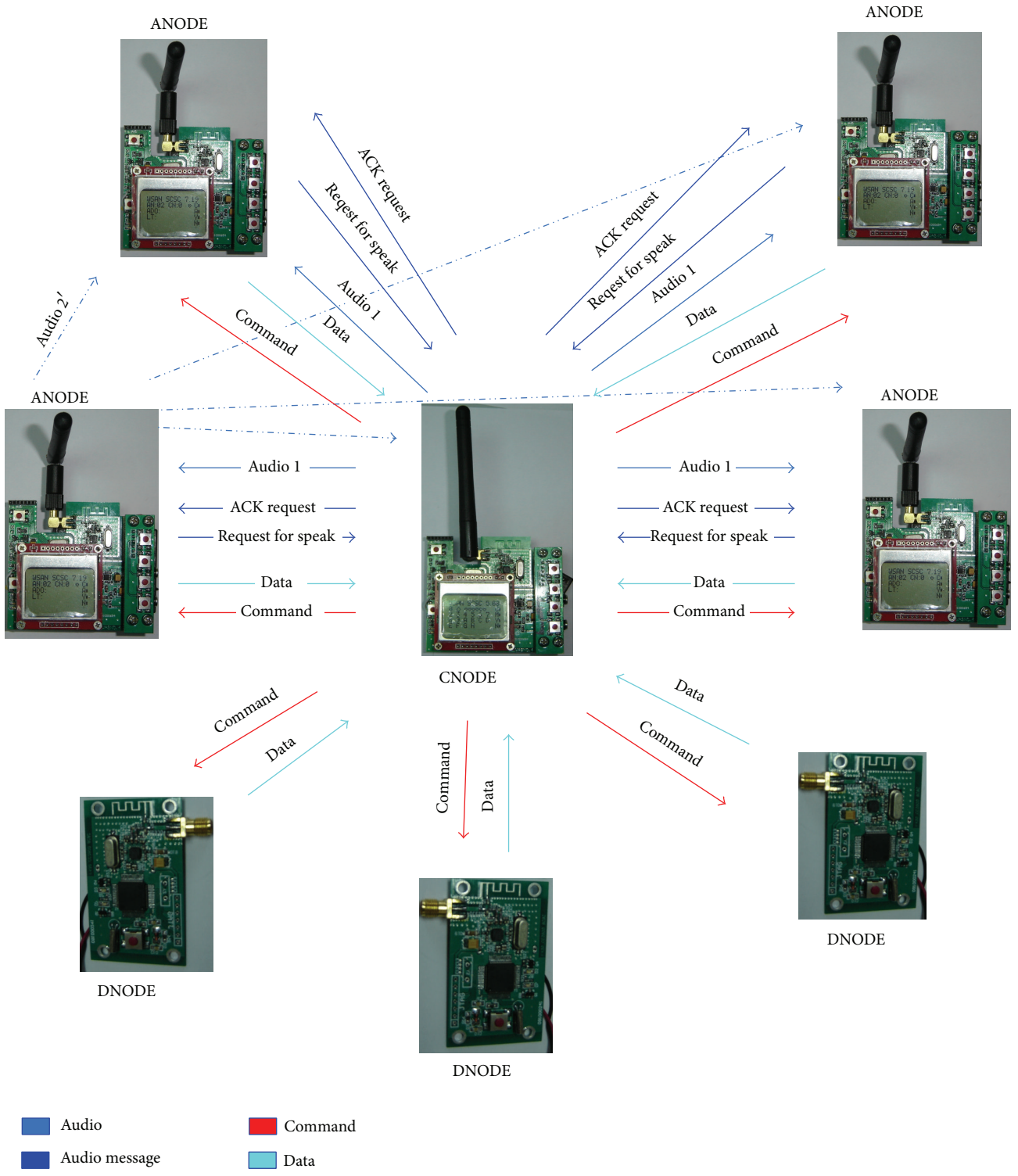


Figure 18: Schematic diagram of VCF.

References

[1] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Computer Networks*, vol. 52, no. 12, pp. 2292–2330, 2008.

[2] S. A. Khan, *Wireless Sensor Networks: Current Status and Future Trends*, CRC Press, New York, NY, USA, 2012.

[3] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, "A survey on wireless multimedia sensor networks," *Computer Networks*, vol. 51, no. 4, pp. 921–960, 2007.

[4] I. T. Almalkawi, M. G. Zapata, J. N. al-Karaki, and J. Morillo-Pozo, "Wireless multimedia sensor networks: current trends and future directions," *Sensors*, vol. 10, no. 7, pp. 6662–6717, 2010.

[5] S. Pudlewski, T. Melodia, and A. Prasanna, "Compressed-sensing-enabled video streaming for wireless multimedia sensor networks," *IEEE Transactions on Mobile Computing*, vol. 11, no. 6, pp. 1060–1072, 2012.

[6] H. Touil, Y. Fakhri, and M. Benattou, "Energy-efficient MAC protocol based on IEEE 802. 11 e for Wireless Multimedia Sensor Networks," in *Proceedings of the International Conference on Multimedia Computing and Systems (ICMCS '12)*, pp. 53–58, 2012.

[7] Y. Zhenyu, L. Ming, and L. Wenjing, "CodePlay: live multimedia streaming in VANETs using symbol-level network coding," *IEEE Transactions on Wireless Communications*, vol. 11, pp. 3006–3013, 2012.

[8] L. L. Hanzo, C. Somerville, and J. Woodard, *Voice and Audio Compression for Wireless Communications*, Wiley-IEEE Press, New York, NY, USA, 2008.

[9] I. F. Akyildiz, T. Melodia, and K. R. Chowdury, "Wireless multimedia sensor networks: a survey," *IEEE Wireless Communications*, vol. 14, no. 6, pp. 32–39, 2007.

[10] V. Gabale, B. Raman, K. Chebrolu, and P. Kulkarni, "LiT MAC: Addressing the challenges of effective voice communication in a low cost, low power wireless mesh network," in *Proceedings of the1st ACM Symposium on Computing for Development (DEV '10)*, p. 5, December 2010.

[11] Q. Li, M. Zhang, and G. Xu, "A novel element detection method in audio sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 607187, 12 pages, 2013.

[12] E. Touloupis, A. Meliones, and S. Apostolacos, "Speech codecs for high-quality voice over ZigBee applications: evaluation and implementation challenges," *IEEE Communications Magazine*, vol. 50, no. 4, pp. 122–128, 2012.

[13] G. Zhao, H. Ma, Y. Sun, and H. Luo, "Design and implementation of enhanced surveillance platform with low-power wireless audio sensor network," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 854325, 18 pages, 2012.

[14] S. W. Jung, H. J. Lee, J. H. Lee, and S. H. Cho, "Interworking of Voice over Sensor Network (VoSN) using the TDMA/TDD MAC and VoIP based SIP," in *Proceedings of the 3rd IEEE International Conference on Network Infrastructure and Digital Content (IC-NIDC '12)*, pp. 97–101, 2012.

[15] V. Gabale, J. Patani, R. Mehta, R. Kalyanaraman, and B. Raman, "Building a low cost low power wireless network to enable voice communication in developing regions," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 16, pp. 2–15, 2012.

[16] J. M. Valin, "Speex: a free codec for free speech," in *Proceedings of the Australian National Linux Conference*, Dunedin, New Zealand, 2006.

[17] J. Kim, J. H. Yoo, J. H. Lee, and S. H. Cho, "A design of the full-duplex voice mixer for multi-user voice over sensor networks (VoSN) systems," in *Proceedings of the 3rd IEEE International Conference on Network Infrastructure and Digital Content (IC-NIDC '12)*, pp. 102–105, 2012.

[18] J. H. Lee, J. H. Yoo, and S. H. Cho, "A new TDMA/TDD MAC protocol design for voice communications based on the IEEE 802.15.4 PHY," in *Proceedings of the 27th International Technical Conference on Circuits/Systems, Computers and Communications*, pp. 1–4, 2012.

[19] L. Herrera, A. Calveras, and M. Catalán, "A two-way radio communication across a multi-hop wireless sensor network based on a commercial IEEE 802.15. 4 compliant platform," *Procedia Engineering*, vol. 25, pp. 1045–1048, 2011.

[20] O. O. Khalifa, S. Khan, M. D. R. Islam, M. B. Muktar, and Z. Yaacob, "Speech coding for bluetooth with CVSD algorithm," in *Proceedings of the RF and Microwave Conference (RFM '04)*, pp. 227–229, October 2004.

[21] CML Microcircuits, "CMX649 Datasheet," http://pdf.datasheetcatalog.com/datasheets/480/101671_DS.pdf.

[22] W. Ye, J. Heidemann, and D. Estrin, "An energy-efficient MAC protocol for wireless sensor networks," in *Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '02)*, pp. 1567–1576, June 2002.

[23] Y. Li, C. S. Chen, Y. Q. Song, and Z. Wang, "Real-time QoS support in wireless sensor networks: a survey," in *Proceedings of the 7th IFAC International Conference on Fieldbuses & Networks in Industrial & Embedded Systems (FeT '07)*, 2007.

[24] B. Sundararaman, U. Buy, and A. D. Kshemkalyani, "Clock synchronization for wireless sensor networks: a survey," *Ad Hoc Networks*, vol. 3, no. 3, pp. 281–323, 2005.

Hindawi Publishing Corporation
Journal of Electrical and Computer Engineering
Volume 2014, Article ID 910658, 14 pages
http://dx.doi.org/10.1155/2014/910658

Research Article

Evaluation of Power Saving and Feasibility Study of Migrations Solutions in a Virtual Router Network

V. Eramo, S. Testa, and E. Miucci

Department of Information Engineering, Electronics and Telecommunications (DIET), Sapienza University of Rome, 00184 Rome, Italy

Correspondence should be addressed to V. Eramo; vincenzo.eramo@uniroma1.it

Received 9 November 2013; Accepted 8 January 2014; Published 30 March 2014

Academic Editor: Yan Luo

The power consumption of the network equipment has increased significantly and some strategies to contain the power used in the IP network are needed. Among the green networking strategies, the virtualization class and in particular the deployment of migrating virtual routers can lead to a high energy saving. It consists in migrating virtual routers in fewer physical nodes when the traffic decreases allowing for a power consumption saving. In this paper we formulate the problem of minimizing the power consumption as a Mixed Integer Linear Programming (MILP) problem. Due to the hard complexity of the introduced MILP problem, we propose a heuristic for the migration of virtual routers among physical devices in order to turn off as many nodes as possible and save power according to the compliance with network node and link capacity constraints. We show that 50% of nodes may be turned off in the case of a real provider network when traffic percentage reduction of 80% occurs. Finally we also perform a feasibility study by means of an experimental test-bed to evaluate migration time of a routing plane based on QUAGGA routing software.

1. Introduction

The number of Internet users today is over 2 billion and it will increase in the next few years. This will lead to a significant growth of energy consumption; just consider that in 2012 the Internet energy consumption has been the 9,8% of the overall energy consumption in the USA [1]. All the strategies proposed in the literature to reduce the power consumption of Internet or to improve the utilization of the network devices already deployed (nodes, links) are indicated as Green Networking. Within this big area, four classes of strategies can be identified according to the type of network equipment considered for reducing the power consumption and the mechanism used to obtain this gain: resource consolidation, selective connectedness, proportional computing, and virtualization [1]. The resource consolidation class is related to the dimensioning strategies to reduce the global consumption due to devices underutilized at a given time. The selective connectedness class regroups distributed mechanism allowing single pieces of equipment to go idle for some time, as transparently as possible to the rest of the network devices [2–5]. The proportional computing category, introduced in [6], starts from the assumption that a device can exhibit different energy consumption profiles as a function of its utilization level and this energy-aware profiles offer different optimization opportunities. The Virtualization class takes into account the mechanisms allowing more than one service to operate on the same hardware improving its utilization [7, 8].

We have to notice that a huge part of the energy in a network is used to supply switching nodes and data center [1]. Then it is also important to underline the fluctuating pattern of the traffic in the network in a day: we observe traffic peak during the day time and large reductions during the night time. Starting from these two assumptions and focusing on the virtualization class features we are interested in proposing and evaluating solutions based on router virtualization and migration that allow for a reduction of the power consumption in Internet networks. The proposed solutions employ the traffic daily variations: as the network traffic volume decreases in the night, virtual routers can be migrated to a smaller set of physical routers and the unneeded

physical routers can be shut down or put into hibernation mode to save power. Obviously this physical node sharing mechanism is largely known in literature as consolidation [1, 9, 10] and it is inspired by virtual machine consolidation allowing for energy saving in cloud datacenter environment [11–14]. Recently energy-aware virtual network embedding through consolidation has been studied [15–17]. These few studies propose solutions to reduce energy consumption in environments in which virtual networks are embedded in a shared substrate run by an infrastructure provider. In this paper we study the case in which one only virtual router layer is hosted on an MPLS/IP network supporting the virtual router migration. When the traffic decreases, some virtual routers can migrate towards other MPLS/IP nodes by employing the reconfiguration capacity of the MPLS layer. The problem of choosing the migrating virtual routers and the MPLS/IP nodes to host them can be modeled as an optimization problem, but due to its high complexity, we introduce a heuristic to evaluate the power saving and the effectiveness of the virtual router migration technique. We also study the impact that the virtual router migration has on the MPLS layer in terms of number of label switched paths (LSP) to be reconfigured.

The rest of the paper is organized as follows. In Section 2 we present the main technical features to implement the virtual router migration (VRM). A Mixed Integer Linear Programming (MILP) formulation of the power consumption minimization problem in a virtual router network is given in Section 3. In Section 4 we describe the proposed heuristic. Some simulation results are shown in Section 5. A feasibility study is carried out in Section 6 where we describe an experimental test-bed to evaluate the migration time of QUAGGA-based routing plane. The conclusions and future research items are finally reported in Section 7.

2. Virtual Router Migration Technique

We start considering the Virtual ROuters On the Move (VROOM) paradigm proposed in [7], in which the assumption, confirmed in practice, is that a logical router instance can migrate among physical nodes. Clearly the migration cannot take place without verifying some constraints; in particular before and after the migration all the logical configurations or states must remain the same. Therefore the IP addresses must remain the same as well as the routing protocol configurations and the overall logical topology; furthermore we want to avoid EGP/IGP reconvergence and routing protocol adjacencies loss. The possible applications of this paradigm are different; for example, it can be used to perform planned maintenance without service disruption. In our contest we want to use it for power saving strategies: when the traffic decreases significantly we want to move the virtual routers from a physical device to another in order to turn off the first machine and save power. An example of migration is reported in Figure 1. Seven Physical Elements (PE) and seven virtual routers (VR) are shown. Initially a VR is located in each PE as indicated in Figure 1(a). The migration of two VRs allows us to switch off two PEs as indicated in Figure 1(b) and to save in network power consumption.

There are three main aspects that make possible the VROOM paradigm: first of all the possibility of creating virtual router instances, then a clear separation between control and data plane, and finally the dynamic binding of virtual interfaces on the physical ones. The migration process starts creating a copy of all the information about the control plane of the virtual router (link state database, network interfaces configuration, etc.), then they are sent to the new physical device and here the data plane is cloned, finally the virtual interfaces are mapped on the new physical ones. Another technological aspect that makes possible the migration of virtual router instances is the deployment of reconfigurable Transport Network, that is to say, a protocol layer between IP and physical layer able to easily reconfigure the paths of IP flows. For example, MPLS or an optical network with its lightpaths can be seen as reconfigurable transport network. An IP/MPLS network is shown in Figure 2 before (a) and after (b) the migration of a virtual router. On the top of IP physical routers (PHYs) the virtual routers (VRs) connected by virtual links (L) are depicted. Each virtual link is mapped at the MPLS level on a label switched path (LSP), that is, a list of connected MPLS routers. For example, when the VR-1 hosted by the physical node PHY-A migrates to PHY-B the logical links L2 and L3 need to be remapped using the MPLS substrate: in particular the corresponding LSP-2 and LSP-3 change their path. After migration the physical node PHY-A which remained idle is turned off.

3. A Mixed Integer Linear Programming Formulation of the Migration Problem to Minimize the Power Consumption

For our research we have considered an IP/MPLS network on which a virtual network is mapped. Each IP physical router hosts a virtual router and each virtual link is mapped on a label switched path (LSP) of the MPLS substrate network. In this scenario, when a migration node has to be moved, the reconfigurability of the MPLS network is exploited to displace all of the LSPs on which the virtual links are mapped so that the overall virtual network topology remains the same.

The virtual nodes have to be moved so that the power consumption is minimized and both link bandwidth and node processing capacity are available to move virtual links and virtual router, respectively. Next we adapt the Mixed Integer Linear Programming (MILP) formulation proposed in [17] for our migration problem.

(1) Notation and Parameters

(i) N: number of virtual routers/physical nodes;
(ii) PHY_n $(n = 1, \ldots, N)$: nth physical nodes;
(iii) VR_i $(i = 1, \ldots, N)$: ith virtual router;
(iv) $b_{h,k}$ $(h, k = 1, \ldots, N)$: offered traffic between VR_h and VR_k;
(v) $\Theta_{\mathrm{PHY},i}$ $(i = 1, \ldots, N)$: set of physical nodes in which VR_i may migrate;

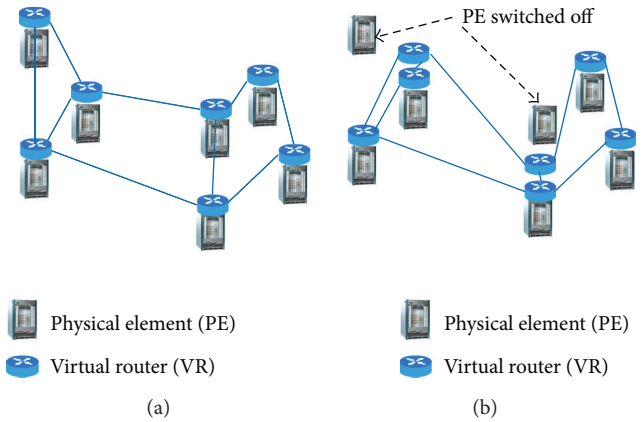


FIGURE 1: An example of virtual router migration for reducing the power consumption in Internet. The network is composed of seven physical elements (PE) and seven virtual routers (VR) (a). Two VRs migrated and two PEs can be switched off (b).

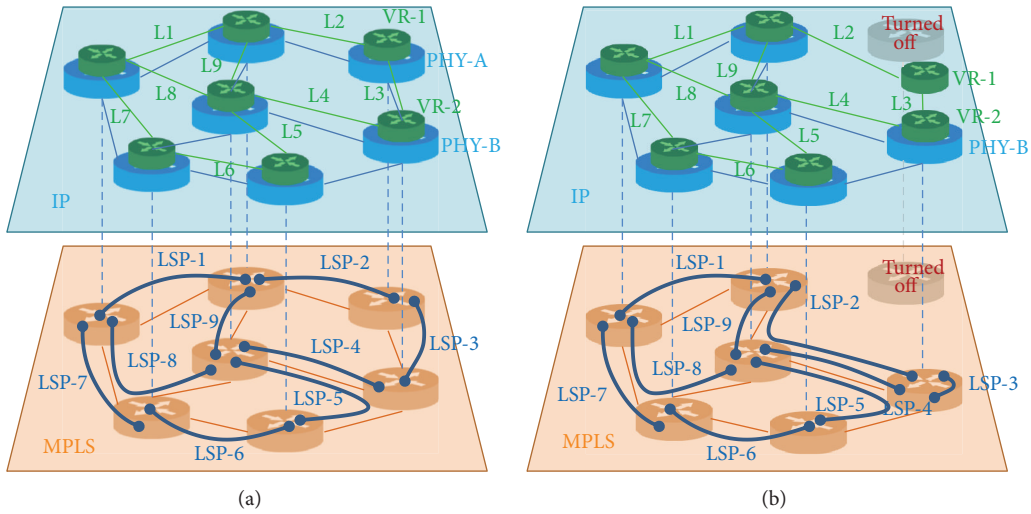


FIGURE 2: IP/MPLS network configuration before (a) and after (b) the migration of a virtual router.

(vi) $\lambda_{\mathrm{PHY}_n}^{\max}$ $(n = 1, \ldots, N)$: maximum traffic that PHY_n can handle;

(vii) P_{PHY_n} $(n = 1, \ldots, N)$: consumption power of the physical node PHY_n; it takes into account mainly the power consumed by the chassis and the route processor;

(viii) $P_{i,j}^{\mathrm{link}}$ $(i, j = 1, \ldots, N)$: power consumed by the Line Card in PHY_i connecting PHY_i to PHY_j;

(ix) $d_{\max}$: maximum degree of the nodes in the physical network;

(x) $c_{i,j}$ $(i, j = 1, \ldots, N)$: link bandwidth if P_{PHY_i} and P_{PHY_j} are connected;

(xi) β: link capacity overprovisioning factor.

(2) Variables

(i) α_i^j $(i, j = 1, \ldots, N)$: binary variable assuming the value 1 if VR_i is allocated to PHY_j; otherwise its value is zero;

(ii) γ_i $(i = 1, \ldots, N)$: binary variable assuming the value 1 if PHY_i is switched on after the migration; otherwise its value is zero;

(iii) $\delta_{i,j}$ $(i, j = 1, \ldots, N)$: binary variable assuming the value 1 if the link connecting PHY_i and PHY_j is switched on after the migration;

(iv) $z_{h,k}^{i,j} = \alpha_h^i \alpha_k^j$ $(i, h, j, k = 1, \ldots, N)$: binary variable introduced to make linear the optimization problem;

(v) t_i^j $(i, j = 1, \ldots, N)$: amount of bandwidth allocated between PHY_i and PHY_j to support the bandwidth demand of the virtual routers mapped on them

$$t_{i,j} = \sum_{h=1}^{N} \sum_{k=1}^{N} b_{h,k} \alpha_h^i \alpha_k^j; \quad (1)$$

(vi) $f_{h,k}^{i,j}$ $(i,h,j,k = 1,\ldots,N)$: amount of bandwidth allocated between PHY_h and PHY_k and passing through the link connecting PHY_i and PHY_j.

(3) Constraints

(i) A constraint introduces the variables $z_{i,h}^{j,k}$ to avoid the nonlinearity on the formulation. Then expression (1) becomes

$$t_{i,j} = \sum_{h=1}^{N}\sum_{k=1}^{N} b_{h,k} z_{h,k}^{i,j}. \tag{2}$$

(ii) The following source, destination, and input/output flow conservation constraints hold:

$$\begin{aligned}
&\sum_{h=1\,(h\neq j)}^{N} f_{i,j}^{h,i} - \sum_{h=1\,(h\neq j)}^{N} f_{i,j}^{i,h} = -t_{i,j}, \quad i,j = 1,\ldots,N\\
&\sum_{h=1\,(h\neq j)}^{N} f_{i,j}^{h,j} - \sum_{h=1\,(h\neq j)}^{N} f_{i,j}^{j,h} = t_{i,j}, \quad i,j = 1,\ldots,N\\
&\sum_{i=1}^{N} f_{h,k}^{i,s} - \sum_{j=1}^{N} f_{h,k}^{s,j} = 0, \quad h,k = 1,\ldots,N;\ s\neq h;\ s\neq k.
\end{aligned} \tag{3}$$

(iii) The correlation between the variables $z_{h,k}^{i,j}, \alpha_h^i, \alpha_k^j$ has to be introduced to guarantee $z_{h,k}^{i,j}$ will be 1 only if α_h^i and α_k^j are 1:

$$\begin{aligned}
&\sum_{i=1}^{N} z_{h,k}^{i,j} = \alpha_k^j, \quad j,h,k = 1,\ldots,N\\
&\sum_{j=1}^{N} z_{h,k}^{i,j} = \alpha_h^i, \quad i,h,k = 1,\ldots,N\\
&\alpha_h^i + \alpha_k^j - z_{h,k}^{i,j} \leq 1, \quad i,j,h,k = 1,\ldots,N.
\end{aligned} \tag{4}$$

(iv) The following link and node processing capacity constraints are introduced by the following expressions:

$$\begin{aligned}
&\sum_{h=1}^{N}\sum_{k=1}^{N} f_{h,k}^{i,j} \leq (1-\beta)\, c_{i,j}\delta_{i,j}, \quad i,j = 1,\ldots,N\\
&\sum_{i=1}^{N}\sum_{h=1(h\neq i)}^{N} \alpha_i^j b_{i,h} + \sum_{i=1}^{N}\sum_{h=1(h\neq i)}^{N} \alpha_i^j b_{h,i} \leq \lambda_{\mathrm{PHY}_j}^{\max},\\
&\qquad\qquad j = 1,\ldots,N.
\end{aligned} \tag{5}$$

(v) We assume that a virtual router is constrained to be mapped on one only physical node belonging to the physical node set in which it can migrate:

$$\begin{aligned}
&\sum_{j=1}^{N} \alpha_i^j = 1, \quad i = 1,\ldots,N\\
&\alpha_i^j = 0, \quad j \notin \Theta_{\mathrm{PHY},i}\ \ (i = 1,\ldots,N).
\end{aligned} \tag{6}$$

(vi) Constraints have to be introduced to guarantee that a physical node is active when at least one of its incoming or outgoing link is active:

$$\begin{aligned}
&\sum_{j=1}^{N} \delta_{i,j} + \sum_{j=1}^{N} \delta_{j,i} \geq \gamma_i, \quad i = 1,\ldots,N\\
&\sum_{j=1}^{N} \delta_{i,j} + \sum_{j=1}^{N} \delta_{j,i} \leq 2d_{\max}\gamma_i, \quad i = 1,\ldots,N.
\end{aligned} \tag{7}$$

(vii) Finally we introduce a constraint that guarantees the switching off of both ingoing and outgoing link of a physical node:

$$\delta_{i,j} = \delta_{j,i}, \quad i,j = 1,\ldots,N. \tag{8}$$

(4) Objective

(i) We have to minimize the total power consumption, that is,

$$\min\left(\sum_{i=1}^{N} \gamma_i P_{\mathrm{PHY}_i} + \sum_{i=1}^{N}\sum_{j=1}^{N} \delta_{i,j} P_{i,j}^{\mathrm{link}}\right). \tag{9}$$

4. A Migration Heuristic for Power Saving

The optimization problem we have defined in Section 3 is complex and can be solved in the case of few physical nodes/virtual nodes. For this reason we propose a heuristic, referred to as Maximum Energy Efficiency (MEE) and whose the main steps are reported in Algorithm 1. Let us introduce the following notations:

(i) $\Lambda_{\mathrm{PHY}} = \{\mathrm{PHY}_n, n = 1,\ldots,N\}$: set of all physical nodes;

(ii) $\Lambda_{\mathrm{VR}} = \{\mathrm{VR}_n, n = 1,\ldots,N\}$: set of virtual nodes;

(iii) λ_{VR_n} $(n = 1,\ldots,N)$: total traffic ingoing/outgoing in/from the VR_n;

(iv) $s_{\max} = \max_{i=1,\ldots,N}\mathrm{Card}(\Theta_{\mathrm{PHY},i})$: the maximum of the cardinalities of the sets $\Theta_{\mathrm{PHY},i}$ $(i = 1,\ldots,N)$.

The following variables are also introduced:

(i) λ_{PHY_n}: total traffic incoming/outgoing in/from all the virtual routers hosted in PHY_n;

(ii) Γ_{PHY_n}: set of PHY_n's adjacent nodes in the physical network.

The proposed heuristic is based on turning off less energy efficient nodes and the migration of VRs towards more energy efficient nodes. For this reason the energy efficiency η_{PHY_n} of the physical node PHY_n $(n = 1,\ldots,N)$ is introduced and defined as

$$\eta_{\mathrm{PHY}_n} = \frac{\lambda_{\mathrm{PHY}_n}}{P_{\mathrm{PHY}_n}^{\mathrm{tot}}}, \tag{10}$$

(1) /∗***LSP set up phase***∗/
(2) **map** each virtual router VR_n on the corresponding physical node PHY_n
(3) **set up** LSPs in the MPLS network carrying traffic incoming/outgoing to/from VR_n according to capacity constraints and a mapping policy
(4) /∗***Physical node turning off phase***∗/
(5) **while** $\Lambda_{PHY} \neq \emptyset$ **do**
(6) **find** $PHY_n \mid \eta_{PHY_n} = \min_{PHY_s \in \Lambda_{PHY}} \eta_{PHY_s}$
(7) $\zeta = \Theta_{PHY,n}$
(8) **while** $\zeta \neq \emptyset$ **do**
(9) **find** $PHY_m \mid \eta_{PHY_m} = \max_{PHY_s \in \zeta} \eta_{PHY_s}$
(10) **if** $\left(\lambda_{PHY_m} + \lambda_{PHY_n} \leq \lambda^{\max}_{PHY_m}\right) \wedge$ (LSPs to/from PHY_m carrying the traffic to/from VR_n can be set up according to capacity constraints and a mapping policy) **then**
(11) **map** VR_n into PHY_m
(12) **turn off** PHY_n
(13) **for** each $PHY_s \in \Gamma_{PHY_n}$ **do**
(14) **update** η_{PHY_s}
(15) **end for**
(16) **update** λ_{PHY_m} and η_{PHY_m}
(17) **if** $PHY_m \in \Lambda_{PHY}$ **then**
(18) $\Lambda_{PHY} = \Lambda_{PHY} \setminus \{PHY_m\}$
(19) **end if**
(20) **for** each $VR_s \in \Lambda_{VR}$ **do**
(21) **if** $PHY_n \in \Theta_{PHY,s}$ **then**
(22) $\Theta_{PHY,s} = \Theta_{PHY,s} \setminus \{PHY_n\}$
(23) **end if**
(24) **end for**
(25) **go to line** (30)
(26) **else**
(27) $\zeta = \zeta \setminus \{PHY_m\}$
(28) **end if**
(29) **end while**
(30) $\Lambda_{PHY} = \Lambda_{PHY} \setminus \{PHY_n\}$
(31) **end while**

Algorithm 1: Maximum energy efficiency.

where $P^{\text{tot}}_{PHY_n}$ is the sum of the node power consumption and the active links power consumption of PHY_n; in particular if the physical node has L_n active ingoing/outgoing links, $P^{\text{tot}}_{PHY_n}$ can be expressed as

$$P^{\text{tot}}_{PHY_n} = P^{\text{C,RP}}_{PHY_n} + \sum_{l=1}^{L_n} P^{\text{LC}}_{n,l}, \tag{11}$$

where $P^{\text{C,RP}}_{PHY_n}$ is the chassis and route processor power consumption of PHY_n and $P^{\text{LC}}_{n,l}$ ($l = 1, \ldots, L_n$) is the power consumption of the lth Line Card. We assume that the node power consumption and the active link power consumption are independent of the offered traffic. In such a way we model the power consumption of today's devices that consume a large amount of static power and a very limited amount of power depending on the current load [18, 19]. However the proposed heuristic can be easily extended to the case in which the node power consumption is dependent on the offered traffic [20].

Next we illustrate the main steps of the proposed heuristic. A preliminary step is needed to map the VRs on the corresponding physical nodes (PHY_s) (line 2) as well as the virtual links on the LSPs of the MPLS network according to a mapping strategy (line 3).

When the IP/MPLS network is correctly configured, MEE chooses the physical device with the least energy efficiency (line 6). Then it choses another physical node in which to migrate the virtual router hosted by the selected node. This physical node is chosen among the nodes of the set in which the virtual node can migrate ($\Theta_{PHY,n}$). In particular the set ζ is introduced and containing the physical nodes belonging to set $\Theta_{PHY,n}$ (line 7). These nodes are selected in order decreasing of energy efficiency (line 9) and the algorithm verifies two constraints (line 10). The first one is that the new node has the necessary capacity to manage the incoming/outgoing traffic of the VR migrating on it. The second constraint refers to the possibility of rerouting the LSPs corresponding to the virtual links afferent to the migrating VR: in particular this problem takes into account

the constraint on the residual physical links capacity on which the LSPs have to be mapped. In Section 5 we will show some results about two LSPs rerouting approaches. The first one is based in solving of a Multicommodity Flow (MCF) problem [21] enabling the splitting of LSPs on more than one path; the commodities and the capacities of the MCF problem are the flows ingoing/outgoing the VR to be migrated and the residual capacities of the MPLS network, respectively. The second approach is based on rerouting each LSP one one shortest path.

If these constraints are verified, the migration of the VR in the new PHY node takes place (line 11), the old physical node is turned off (line 12), and the energy efficiency of the adjacent nodes to the turned off node are properly updated (line 14). The amount of traffic λ_{PHY_m} managed by the new PHY node and its energy efficiency η_{PHY_m} are updated considering the contribution of the VR migrated (line 16). To make simple the heuristic, we remove the physical node PHY_m from Λ_{PHY_m} (line 18), guaranteeing that the VRs on it will not migrate towards other physical nodes. By updating the sets $\Theta_{\mathrm{PHY},s}$ ($s = 1, \ldots N$), future migrations towards the turned off node are avoided (line 22). Finally regardless of whether the VR can be moved or not, the relative physical node is removed from Λ_{PHY} (line 30) in order to avoid a too long simulation time. The algorithm stops when all the nodes that can be turned off have been explored.

Next we report a complexity analysis of the proposed algorithm. When the logical links are remapped on the shortest paths by using the Dijkstra algorithm, the complexity can be evaluated according to the following remarks: (i) the proposed heuristic performs N steps in which at each step the least energy efficient physical node, among the switched on ones, is selected; the complexity of these operations is $O(N^2)$; (ii) the physical nodes in which to try migrating a VR are selected in decreasing order of energy efficiency; the complexity of these operations is $O(s^2_{\max})$; (iii) the LSP rerouting is accomplished by using the Dijkstra algorithm whose complexity, when heap binary data structure is used, is $O(M \log N)$, M being the number of links of the MPLS network. According to these remarks the heuristic complexity is $O(N^2 M s^2_{\max} \log N)$ when LSP rerouting based on shortest path is accomplished.

When the proposed heuristic is based on a MCF rerouting, the complexity is polynomial because a linear programming formulation can be given for an MCF problem.

5. Numerical Results

The power consumption saving that the proposed heuristic allows us to obtain has been evaluated when both two-level hierarchical networks and more general provider networks are taken into account. In the first case we consider a network scenario inspired by the real network of an Internet Service Provider whose structure is hierarchical [22]. The network is composed by few core nodes that are highly interconnected by means of high-capacity links. It is also composed of edge nodes that are used to interconnect aggregation nodes to core nodes. The aggregation nodes are the ones to which users are directly connected. A Digital Subscriber Line Access Multiplexer (DSLAM) and an Optical Line Termination (OLT) in PONs are typical examples of aggregation node. Each node is dual-homed; that is, it is connected to the closest pair of edge nodes to guarantee alternate paths in case of failure. We assume that the network is composed of $N = x + x \times y + x \times y \times z$ nodes where x indicates the number of core nodes, y the number of the edge nodes for each core node, and z the number of aggregation nodes for each edge node. The core network has always a ring topology and furthermore we consider a connectivity factor p which gives the probability that two core nodes not adjacent in the ring are connected with a link. In Figure 3 an example of this topology is shown when $x = 6$, $y = 3$, and $z = 4$.

Next we describe the considered traffic model, how the traffic is routed, and how the link capacities are dimensioned. Only the aggregation nodes can be source and destination of traffic which is assumed initially uniformly distributed in the range $[v_{\min}, v_{\max}]$. The offered traffic is scaled by a factor α chosen according to the dimensioning procedure next illustrated. It is routed in the IP network exploiting the shortest paths between each couple source-destination exchanging data. When all the traffic is routed in the network, the aggregated traffic on an IP virtual link is carried by one LSP in the underlying MPLS network connecting directly the end nodes of the IP virtual links. At the beginning we have as many LSPs as the number of virtual links in the IP network. The capacity of the physical link is dimensioned considering the amount of traffic that has been routed through it. In particular we assume that the network links are dimensioned with capacity values belonging to the set $\vartheta \equiv \{C_i \; i = 1, \ldots N\}$ with $C_i < C_{i+1}$ ($i = 1, \ldots, N-1$). The link capacities and the scale factor α are chosen as follows: (i) the highest load link is equipped with the highest capacity C_N and the scale factor α is chosen so that this link carries a traffic equal to $(1-\beta)C_N$, where the factor β is the capacity overprovisioning factor; (ii) each remaining link of the network is dimensioned with the highest capacity value $C \in \vartheta$ such that the scaled traffic carried on the link is smaller than or equal to $(1-\beta)C$.

Next we evaluate the power consumption saving when a traffic reduction occurs and the virtual router migration is performed according to the heuristic introduced in Section 2. We assume that only core or edge nodes can be turned off and, due to the particular topology, when an edge node is turned off, its two adjacent edge nodes cannot be switched off to avoid isolation of the aggregation nodes connected to it. We also assume that migration of a VR can occur only in physical nodes adjacent to the physical nodes in which the VRs is hosted before the migration.

A first set of results is illustrated in Figures 4 and 5 in which we report the average number and the percentage of switched off nodes, respectively, as a function of the traffic percentage reduction γ in a hierarchical network with $x = 6$, $y = 3$, and $z = 4$. Each curve in Figures 4 and 5 is related to a specific value of the connectivity degree p. Each reported value is the mean of values obtained for five traffic realizations. When $\gamma = 0$ the link capacities are dimensioned according to the previous illustrated procedure. The offered traffic between each couple of aggregation nodes

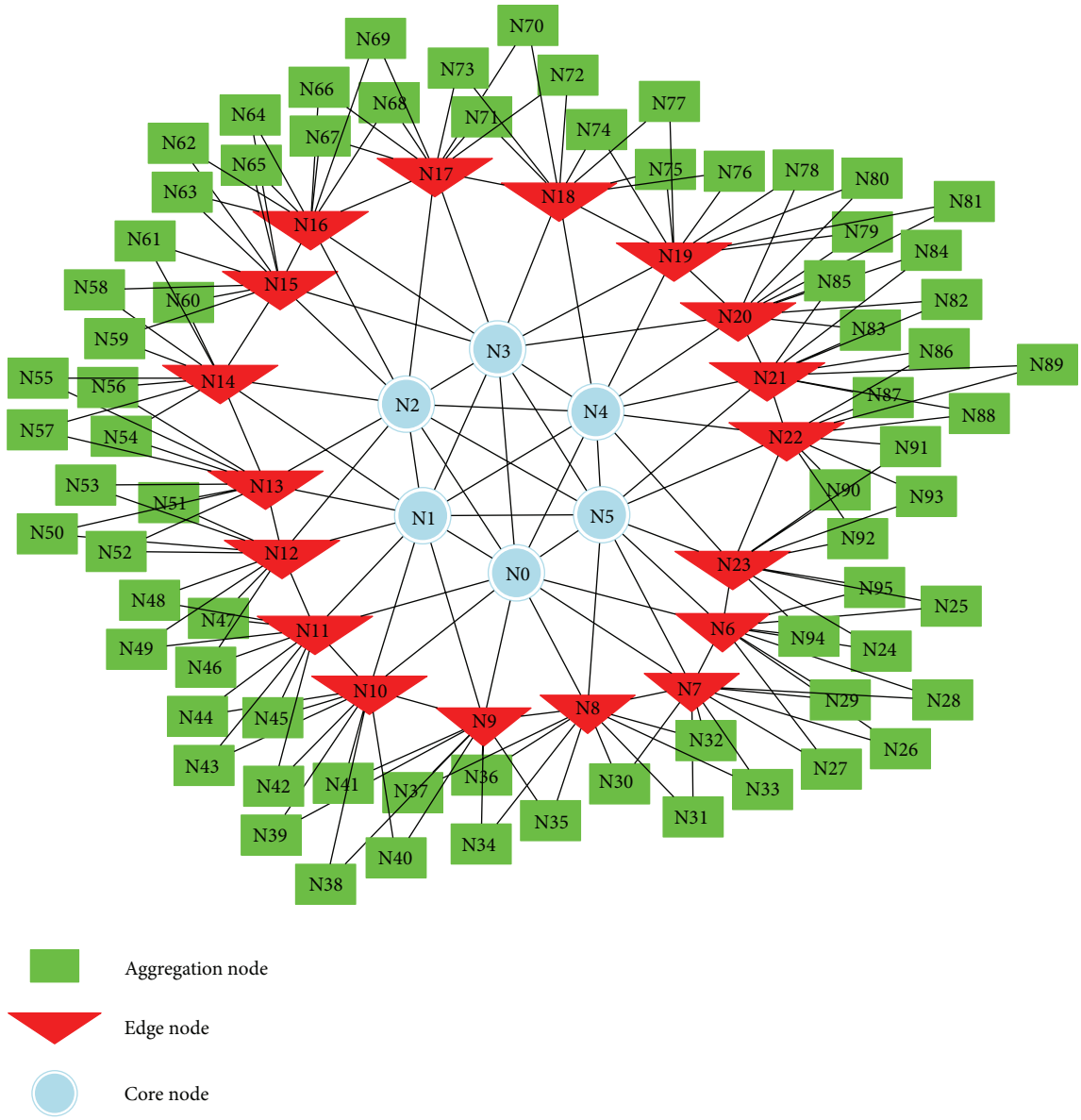


Figure 3: Hierarchical Network topology with $x = 6$, $y = 3$, and $z = 4$.

is characterized by parameter values $v_{\min} = 0{,}5$ and $v_{\max} = 1{,}5$. The following $N = 7$ capacity values are available: $C_1 = 100$ Mb/s, $C_2 = 622$ Mb/s, $C_3 = 1$ Gb/s, $C_4 = 2{,}488$ Gb/s, $C_5 = 2 \times 2{,}488$ Gb/s, $C_6 = 3 \times 2{,}488$ Gb/s, and $C_7 = 4 \times 2{,}488$ Gb/s. The spare capacity is determined by a parameter value $\beta = 0{,}2$. We also assume that when a traffic reduction happens and the virtual router migration procedure is activated, a migration is performed when it is guaranteed that the processing load of each core (edge) node is not higher than the one of the highest load core (edge) node in the traffic condition $\gamma = 0$.

The power consumption values used for the chassis, the route processor, and the Line Cards in edge and core nodes are the ones measured by the authors in [23] and reported in Table 1. We show in Figures 4 and 5 some results that compare two strategies for the remapping of virtual links on new LSPs after the migration of a virtual router. The first one is based on the solution of a Multicommodity Flow (MCF)

Table 1: Values of power consumed by the physical router components.

	Power consumption
Chassis and route processor	
Edge router	220 W
Core router	400 W
Line Cards	
Fast Ethernet	26 W
OC-12 622.08 Mbps	18 W
Gigabit Ethernet	30 W
OC-48 2.488 Gbps	70 W

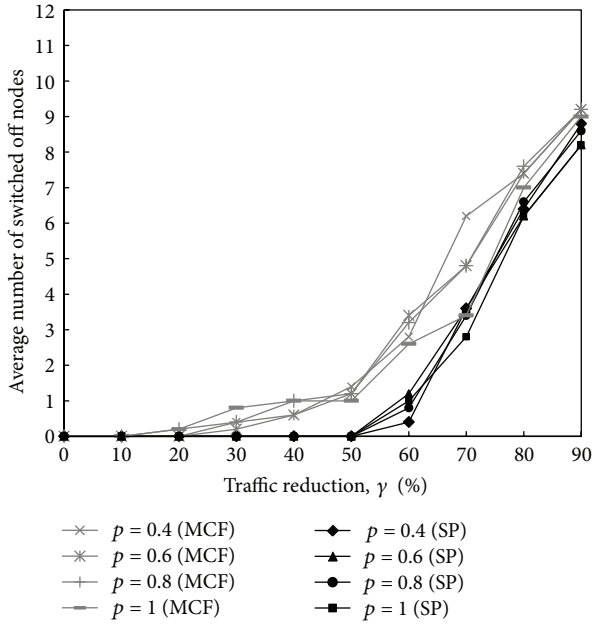


Figure 4: Average number of switched off nodes obtained using MEE Heuristic with MCF and Dijkstra rerouting approach and in the case of the Network 6-3-4.

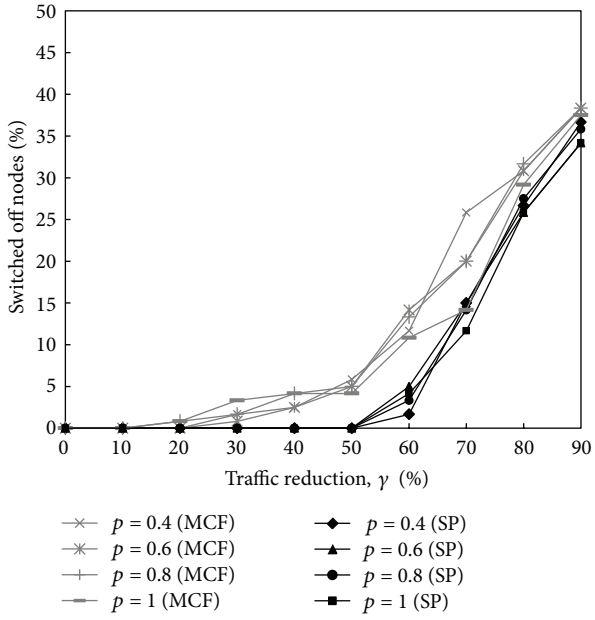


Figure 5: Percentage of switched off nodes obtained using MEE Heuristic with MCF and Dijkstra rerouting approach and in the case of the Network 6-3-4.

problem in order to remap the virtual links; in this case a traffic demand between two virtual routers can be rerouted on more than one path in order to better use the resources on the physical link. In the second strategy it is remapped on the Shortest Path (SP) chosen applying the Dijkstra algorithm. It is easy to see from Figures 4 and 5 that more nodes can be switched off when the traffic decreases. That is obviously due

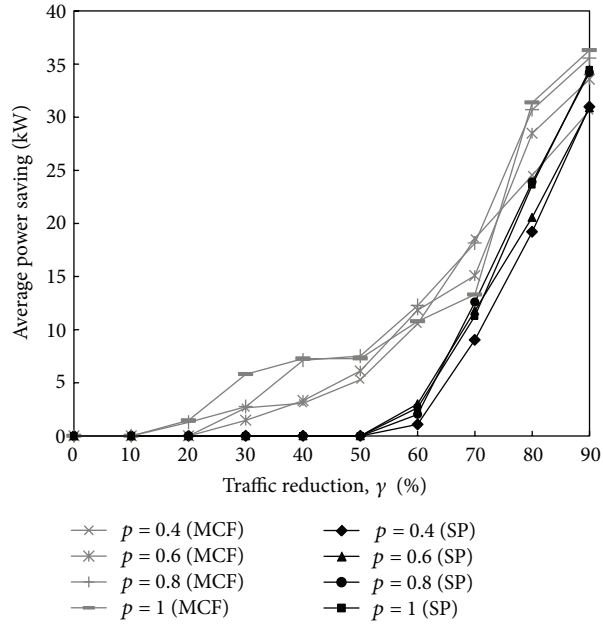


Figure 6: Power consumption saving obtained using MEE Heuristic with MCF and Dijkstra rerouting approach and in the case of the Network 6-3-4.

to the higher availability of resource in the links and nodes. The performance is little dependent on factor p and this is due to the way in which we scale the traffic on the links: the greater the value of p is, the greater the factor α is and this leads to a more connected network but with a higher level of traffic load. The curves related to the shortest path rerouting approach show lower performance and this is due to the fact that in this case a single path is chosen for each LSP to reroute with respect to the case of MCF in which a traffic demand between two virtual routers can be split on more than one path. As a matter of example we notice from Figure 4 that when $\gamma = 80\%$, seven and six nodes can be switched off in the case of rerouting strategies based on MCF and SP, respectively. That leads to switch off the 33% and 27% of nodes as shown in Figure 5. Notice that we evaluate the percentage of switched off nodes only taking into account the nodes possible to switch off.

The power saved turning off the nodes is shown in Figure 6 and the same values in percentage compared with the total power consumption of the overall network is reported in Figure 7.

Finally we report in Figure 8 the average number of LSPs that is necessary to reroute as a result of the shutdown of the nodes. As expected the curves shown in this figure follow the trend of those in Figures 4 and 5 since they are quantities directly dependent. In particular the shortest path rerouting approach, using Dijkstra algorithm, needs to reconfigure a lower number of LSPs but, at the same time, it turns off fewer physical nodes than the MCF rerouting approach. This second strategy allows us to turn off nodes already when the traffic load is around 30% and this is paid with an increasing

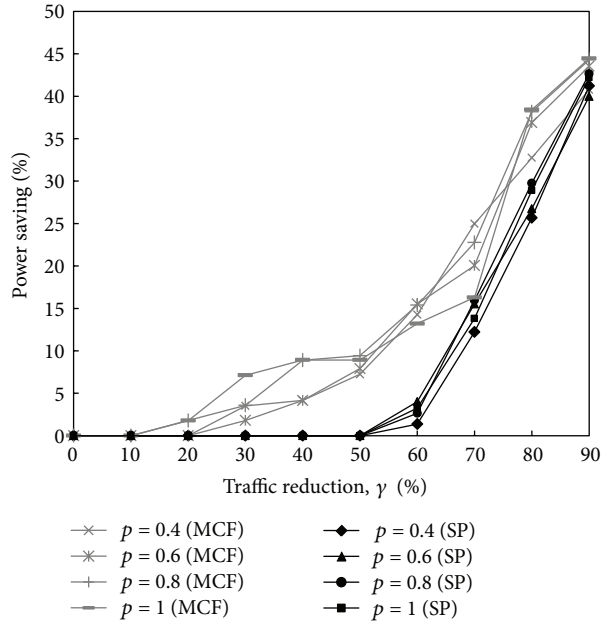


Figure 7: Percentage of power consumption saving obtained using MEE Heuristic with MCF and Dijkstra rerouting approach and in the case of the Network 6-3-4.

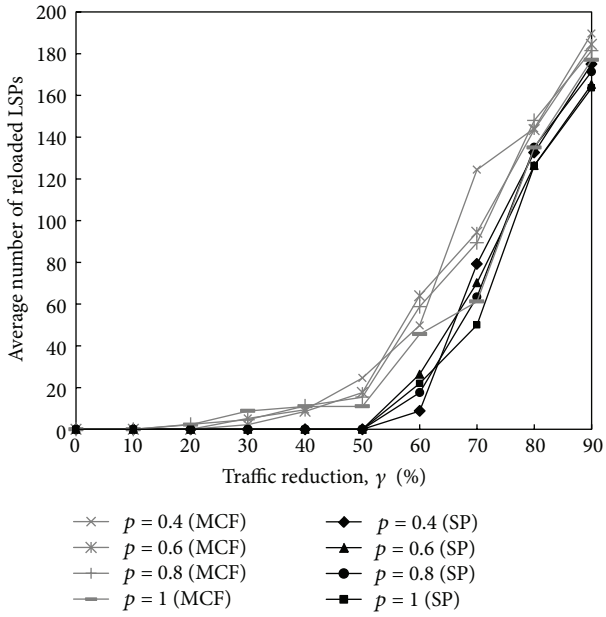


Figure 8: Number of Reloaded LSPs using MEE Heuristic with MCF and Dijkstra rerouting approach for the Network 6-3-4.

number of paths that need reconfiguration in the MPLS domain.

Next we evaluate the power consumption saving in the case of a real provider network. We consider the EBONE network [5] composed of 159 nodes and 614 links and reported in Figure 9(a). We evaluate the power consumption saving when the proposed heuristic with LSP rerouting based on shortest path is applied. We assume that (i) all of the nodes are core nodes; (ii) the constraint that a VR hosted in leaf node cannot migrate; (iii) the constraint that a VR can migrate only in physical nodes adjacent to the physical node in which it is initially hosted; (iv) the traffic is generated among any node couple according to the procedure before described with $v_{\min} = 0{,}5$ and $v_{\max} = 1{,}5$; (v) the availability of Gigabit Ethernet Line Cards and in particular the links can be dimensioned with capacity values equal to nC_{GE} ($n = 1, \ldots, 10$) with $C_{\mathrm{GE}} = 1$ Gb/s.

We report in Figures 9(b), 9(c), and 9(d) and in white color, the nodes that the proposed migration heuristic allows us to obtain in the case of $\gamma = 0\%$, $\gamma = 60\%$, and $\gamma = 80\%$, respectively. Average number and percentage of switched off nodes, power consumption saving and relative percentage, and number of LSPs rerouted are reported in Table 2 for the EBONE network in the case of traffic reduction γ from 0 to 80%. We can notice that a big power saving can be obtained. For instance when $\gamma = 80\%$, as much as 60% of nodes can be switched off.

6. Experimental Test-Bed for the Migration Time Evaluation of a Quagga-Based Routing Plane

We illustrate an experimental test-bed to evaluate the migration time of a routing plane based on QUAGGA routing software. The operation mode correctness of the OSPF routing protocol is also verified. The realized test-bed allows for an evaluation of the migration time as a function of the number of nodes of an emulated network. The section is organized as follows. The software router architecture and the used software are described in Section 6.1. The test-bed realized to evaluate the routing plane migration time is illustrated in Section 6.2 where the main numerical results are also shown.

6.1. Software Router Architecture. There are three main aspects that make possible the virtual router migration paradigm [7]: first of all the possibility to create virtual router instances, then a clear separation between routing and data plane, and finally the dynamic binding of virtual interfaces on the physical ones. A Software Router (SR) equipped with software modules for the implementation of the virtual router migration paradigm is illustrated in Figure 10. It is based on the following main software modules: *Linux* Operating System, *Linux Containers (LXC)* [24] virtualization software, *QUAGGA* [25] routing software, and the *Linux bridge* [26] software. Thanks to the Linux Containers we are able to divide the resources of a Personal Computer (PC) among different virtual routers instances. Each of them has an independent control plane to execute applications, configurations, routing protocols instances, Routing Information Base (RIB), and also an own data plane managing interfaces and Forwarding Information Base (FIB). The isolation of different virtual routers makes possible the migration of one of them transparently with respect to the others. The virtual routers are activated and deactivated by the *VR-Manager*. The virtual interfaces of the virtual routers are mapped on the host physical ones thanks to the *Linux bridge* software that makes

Table 2: Number and percentage of switched off nodes, power consumption saving and relative percentage, and number of LSPs rerouted are reported for the EBONE network in the case of traffic reduction γ from 0 to 80%.

γ	Number of turned off nodes	Percentage of turned off nodes	Power consumption saving (W)	Percentage of power consumption saving	Number of rerouted LSPs
0%	30.8	23%	21332	18%	92.8
20%	50.6	38%	35336	30%	164.8
40%	68.8	52%	49444	42%	280
60%	77.6	59%	58376	50%	364
80%	79.8	60%	60624	52%	384

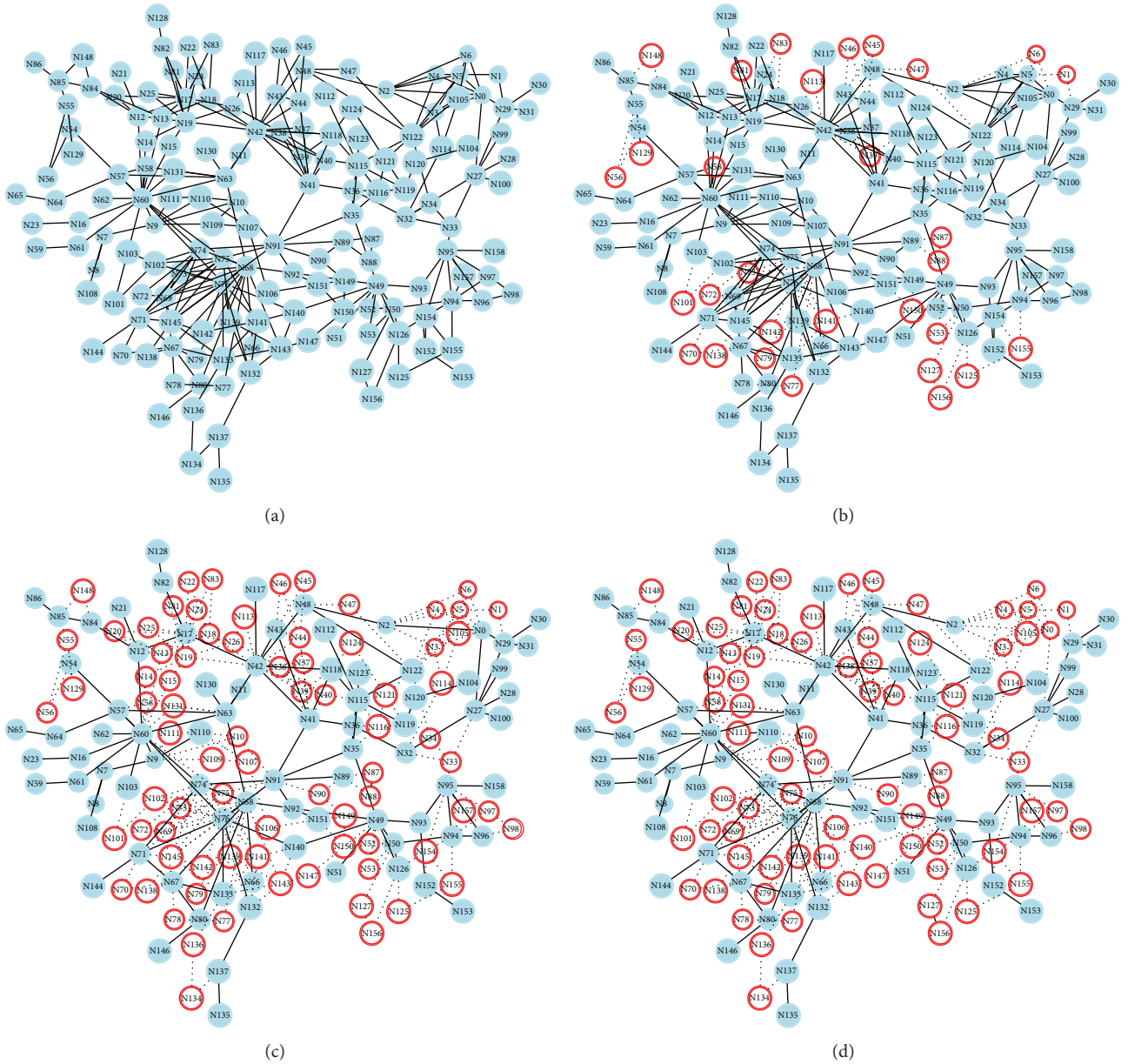


Figure 9: EBONE Network Topology (a); virtual router migration and switching off of physical nodes when $\gamma = 0\%$ (b), $\gamma = 60\%$ (c) and $\gamma = 80\%$ (d), respectively.

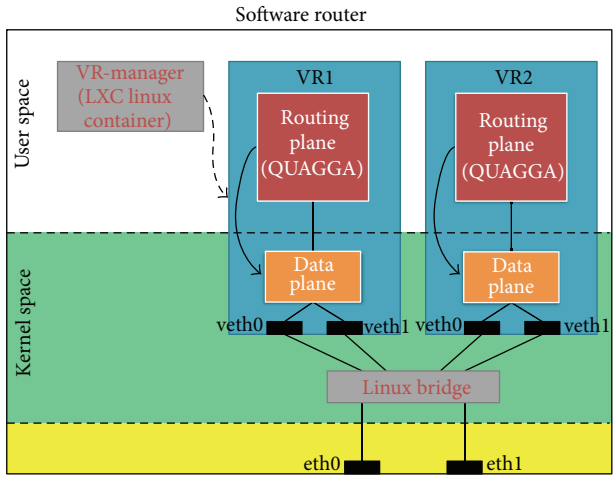


Figure 10: Software Router Architecture supporting virtualization.

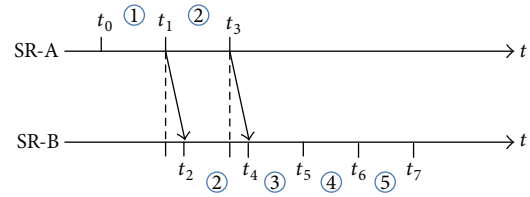


Figure 11: Main steps of virtual router migration between SR-A and SR-B.

possible setting dynamically the association between virtual routers interfaces and physical ones. Each VR is equipped with QUAGGA routing software [25]. To allow the routing plane to migrate, we have realized a Quagga's patch for the OSPF routing plane migration. It is mainly composed of three software modules: the first one copies all of the routing information (LSA database, configuration files, etc.) to be migrated in a text file appropriately encoded; (ii) the second one manages the transferring of the text file containing the routing information between the two SRs in which the migration occurs; the third one decodes the text file and recreates the OSPF data structure in the VR in which the migration occurred.

The key instants of virtual router migration process from a SR-A to a SR-B are shown in Figure 11. When the migration event starts a virtual container is opened in the destination SR. In the interval $[t_0, t_1]$ all of the routing information is copied in the text file. With the aim to transfer the minimum amount of information, we establish that all the virtual routers in the network have the same Quagga's executable files. For this reason only the information contained in both the Quagga's configuration file and the Link State Database (LSDB) data structure containing the Link State Advertisements (LSA) is appropriately encoded and stored in the text file to be transferred from SR-A to SR-B. This phase takes a time depending on the network dimension: as the number of nodes increases, the number of LSAs and thus the dimension of the LSDB increases, so the transferring of this data will take a longer time when the network dimension increases. During the copy of the LSDB it is necessary that no other subprocess accesses the memory of LSDB to modify it, so in this step it is frozen and the update LSAs that the VR may receive are discarded. In the interval $[t_1, t_3]$ ($[t_2, t_4]$ in SR-B) the text file is transferred from SR-A to SR-B. In this phase the update LSAs received are stored in the LSDB in SR-A but are not acknowledged; in such a way the neighbour router which has generated the update LSA is induced, according to the OSPF protocol operation mode, to retransmit it until the migration process is completed. Then SR-B receives the update LSA when retransmissions occur. In the interval $[t_4, t_5]$, the text file received by SR-B is decoded and from this information both the Quagga's configuration file and the LSDB data structure are rebuilt. In this interval the shortest path tree is calculated starting from the network information collected in the LSDB. At this point the Quagga's zebra demon clones the VR's data plane in SR-B and updates the routing table. Also in these two last intervals VR in SR-A goes on to update the LSDB, without acknowledging any LSA received. Finally in the interval $[t_6, t_7]$, IP addresses are assigned to the virtual interfaces of the VR and these last ones are mapped on the SR-B ones thanks to the Linux Bridge software.

6.2. Test-Bed for the Evaluation of the Routing Plane Migration Time. The realized test-bed is shown in Figure 12 and is composed of a DELL PowerConnect 5524 switch with 24 ports GbE and three PC DELL Optiplex 990 with the following hardware features:

(i) CPU Intel Core i7 2600 @ 3.40 GHz,

(ii) SDRAM DDR3 8 GB,

(iii) Hard disk SATA II 1TB, 7200 rpm,

(iv) 2 Network interfaces (only one used for measures).

Each PC is equipped with Linux Ubuntu 10.10 32-bit OS (kernel version 2.6.38). The SR-A and SR-B are the ones involved in the migration process. VR-A and VR-B are executing in SR-A and SR-B, respectively. The testing PC has the function to generate routing and data traffic. The first one allows for the emulation of any network topology in the SR-A and SR-B. The second one allows the migration process to be disturbed by injecting traffic in the network links.

The following software tools have been used.

(i) *LSA Generator* [27] is an open source software installed in the testing PC. It can generate and send all types of LSA defined within the OSPF protocol. A text file, describing the network topology, is given as input to LSA generator that by injecting appropriate LSAs towards SR-A and SR-B allows any network topology to be emulated in them.

(ii) *RUDE* (*Real-time UDP Data Emitter* [28]) is an open source UDP traffic generator in which the packet rate and dimension can be set. The generated traffic is analyzed in another node of the network using the **CRUDE** (*Collector for RUDE* [28]) software.

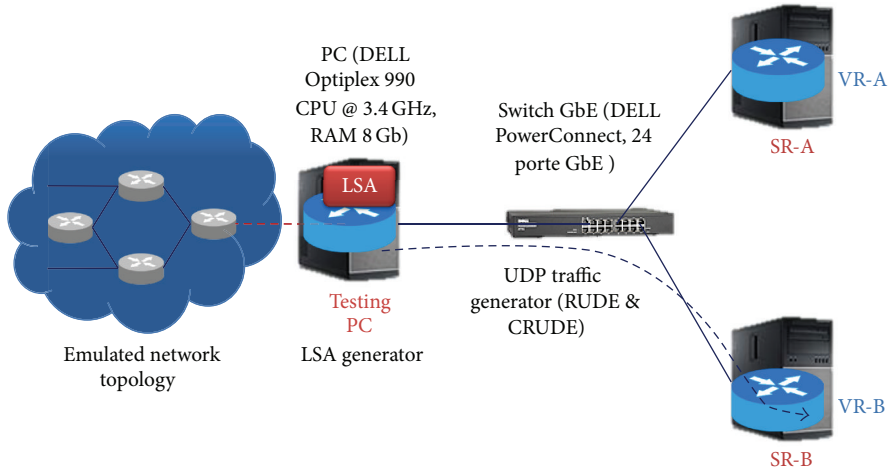


Figure 12: Test-bed for the evaluation of the migration time.

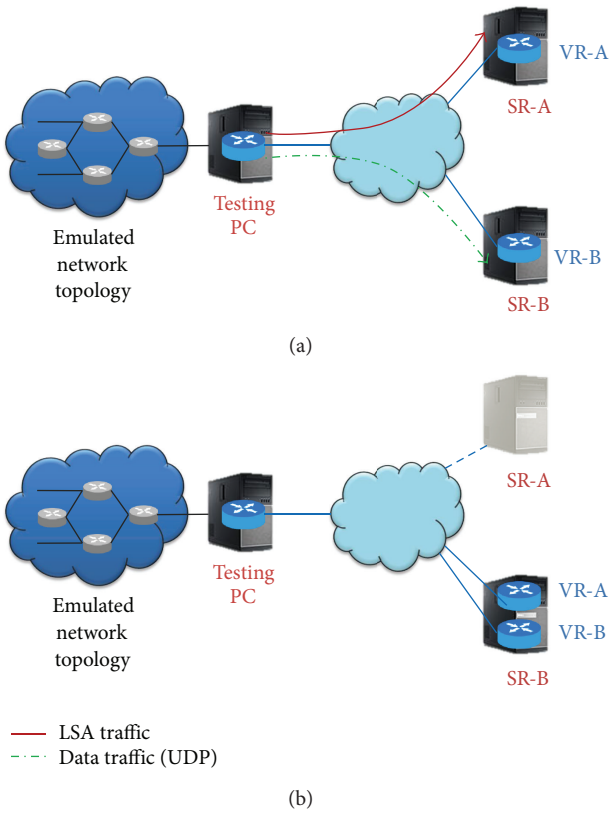


Figure 13: Test-bed scenario before (a) and after (b) the migration of VR-A from SR-A to SR-B.

(iii) *Tshark* [29] monitors and analyses the OSPF routing traffic incoming in the different virtual routers we consider in the simulation.

By means of the introduced test-bed, we have evaluated the Quagga's routing plane migration time and we have verified the operation mode correctness of the OSPF routing protocol when a migration occurs. The test is performed in two phases.

(i) Phase-1. As indicated in Figure 13(a), the testing PC generates the emulated network topology sending LSA towards the SR-A so that the SR-A's LSDB is updated. All of the performed tests are based on fully meshed router networks with each router connected to each other through a different transit network [30]. Hence the testing PC sends a Router LSA and a Network LSA for each router and transit network of the emulated network topology, respectively. It also sends UDP traffic towards the SR-B so that the migration process is disturbed.

(ii) Phase-2. When the emulated network topology is acquired by SR-A, the virtual router VR-A is moved from SR-A to SR-B as shown in Figure 13(b). We have evaluated the various components of the migration time by inserting timers in the Quagga's source codes. The test has been repeated several times varying the number N of nodes (routers and transit networks) from 50 to 500 and then varying the bit rate f_{UDP} of UDP traffic from 1 Mbps up to 700 Mbps, carried in links with 1 Gbps capacity. Finally in this phase we have also verified the operation mode correctness of the OSPF routing protocol.

The migration time is shown in Figure 14 as a function of the number N of nodes of the emulated OSPF network. Moreover the figure provides the information related to the time required by the different migration steps: copy of the routing plane in SR-A ($t_1 - t_0$), transfer of encoded data from SR-A to SR-B ($t_4 - t_2$), and installation on the new physical device SR-B ($t_6 - t_4$) of the routing plane. In this first case study the UDP data traffic rate f_{UDP} is equal to zero. As we can observe the overall migration time grows linearly with the number N of nodes of the emulated network topology and thus with the amount of data of the routing plane that is necessary to transfer. From the results we can notice low

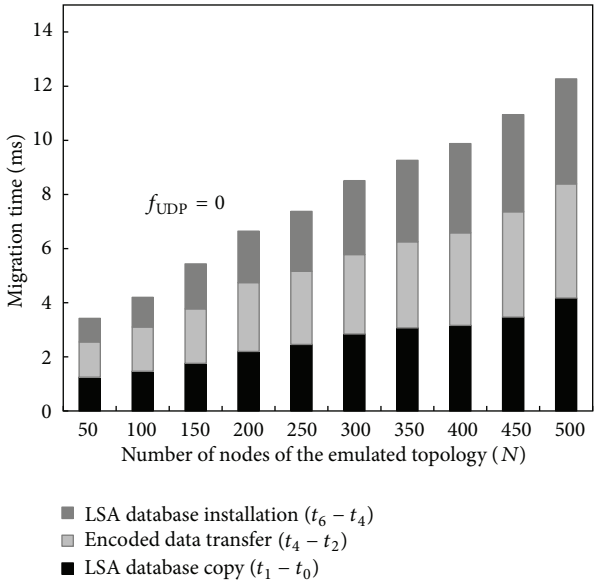


Figure 14: Migration time components as a function of the number N of nodes in the emulated network. The UDP traffic rate f_{UDP} is equal to zero.

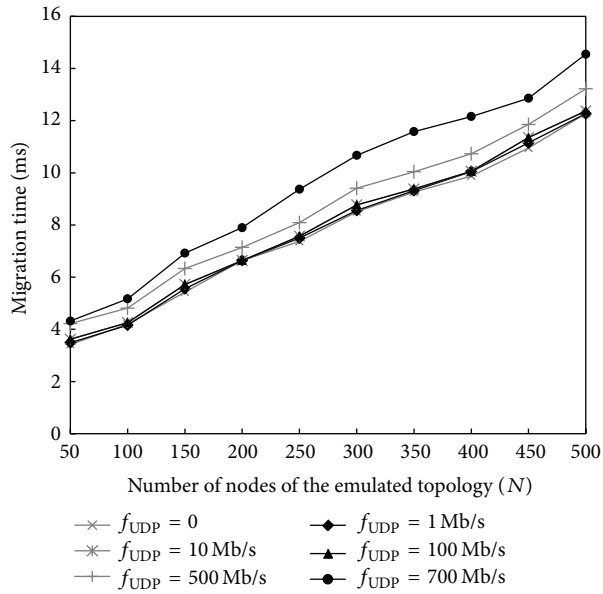


Figure 15: Migration time as a function of the number N of nodes in the emulated network. The UDP traffic rate f_{UDP} is varied from 0 to 700 Mb/s.

migration time that, even in the case of network with $N = 500$ nodes, is smaller than 12 ms.

The migration time is shown in Figure 15 varying the number N of nodes in the emulated network topology, when we consider the UDP data flow sent from the testing PC to the SR-B. Increasing the bit rate, the UDP packets flow tends to saturate the link capacity ingoing the SR-B and this leads to a growing delay in the transferring component $(t_4 - t_3)$ of the routing information from SR-A to SR-B. Even the installation delay component $(t_6 - t_4)$ of the routing plane is increased due to the fact that the SR-B's CPU has to process the UDP packets received. As we can expect the whole migration time grows when the link is stressed by the UDP traffic. For example when the link ingoing the SR-B is 70% loaded, the migration time is increased by 20%.

We have also verified if the migrant virtual router maintains, during the migration, the adjacencies with the testing PC. In fact according to the OSPF protocol, an IP router sends HELLO packets to inform its neighbours that it is active. Two timeouts are used. The first one is the HELLO timeout that establishes the time distance between HELLO packets. The second one is the DEAD timeout that represents the time after which a neighbour is declared not reachable if HELLO packets are not received. In our tests we have considered a HELLO timeout equal to 10, 5, 2, and 1 second and a DEAD timeout set to 4 times the HELLO timeout. We have verified that the VR-A never loses the adjacencies with the testing PC and the VR-B. That is due to the fact that the overall migration time is much lower than the DEAD timeout. Due to the lack of a retransmission mechanism of update LSAs in *LSA Generator*, we are not able to verify the operation mode correctness of the OSPF routing protocol when a topological change occurs and update LSAs are lost. The only consideration we can do is that the very low migration times leads to a low probability that an update LSA is lost.

7. Conclusion

In this paper a heuristic, called Maximum Energy Efficiency (MEE), for virtual router migration in an IP/MPLS network is proposed. The effectiveness of the MEE heuristic has been evaluated in a hierarchical network with core, edge, and aggregation nodes. The simulation results show that the 33% and 27% of nodes can be switched off when the traffic is reduced of 80% in the case of rerouting strategies based on MCF and Dijkstra, respectively. This higher power saving is paid with a higher number of LSPs to be reroute on the MPLS network. We have also evaluated the power consumption saving in real provider networks. In particular the proposed heuristic has been applied to evaluate the power saving in the EBONE network composed of 159 nodes and 614 links. We have shown that a 60% power saving can be reached for a 80% traffic reduction.

Finally the routing plane migration time of a Software Router equipped with QUAGGA routing software has been evaluated in an experimental test-bed made of three PCs: a testing PC and two other PCs between which a virtual router migration is trigged off. The testing PC is able to emulate a network topology and allows for an evaluation of the migration time as a function of the number of nodes in the emulated network. We have evaluated a migration time less than 12 ms in the case of an emulated network of 500 nodes and when the network links are 70% loaded.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The research leading to these results has received funding by MIUR with PRIN 2009 SFINGI (Software router to Improve Next-Generation Internet).

References

[1] A. P. Bianzino, C. Chaudet, D. Rossi, and J. Rougier, "A survey of green networking research," *IEEE Communications Surveys and Tutorials*, vol. 14, no. 1, pp. 3–20, 2012, First Quarter.

[2] K. J. Christensen, C. Gunaratne, B. Nordman, and A. D. George, "The next frontier for communications networks: power management," *Computer Communications*, vol. 27, no. 18, pp. 1758–1770, 2004.

[3] A. Cianfrani, V. Eramo, M. Listanti, M. Marazza, and E. Vittorini, "An energy saving routing algorithm for a green OSPF protocol," in *Proceedings of the IEEE Conference on Computer Communications Workshops (INFOCOM '10)*, San Diego, Calif, USA, March 2010.

[4] A. Cianfrani, V. Eramo, M. Listanti, and M. Polverini, "An OSPF enhancement for energy saving in IP networks," in *Proceedings of the IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS '11)*, pp. 325–330, Shanghai, China, April 2011.

[5] A. Cianfrani, V. Eramo, M. Listanti, M. Polverini, and A. V. Vasilakos, "An OSPF-integrated routing strategy for QoS-aware energy saving in IP backbone networks," *IEEE Transactions on Network and Service Management*, vol. 9, no. 3, pp. 254–267, 2012.

[6] L. A. Barroso and U. Hölzle, "The case for energy-proportional computing," *IEEE Computer*, vol. 40, no. 12, pp. 33–37, 2007.

[7] Y. Wang, E. Keller, B. Biskeborn, J. van der Merwe, and J. Rexford, "Virtual routers on the move: live router migration as a network-management primitive," in *Proceedings of the ACM Conference on Data Communication (SIGCOMM '08)*, pp. 231–242, Seattle, Wash, USA, August 2008.

[8] X. Chen and C. Philips, "Virtual router migration and infrastructure sleeping for energy management of IP over WDM networks," in *Proceedings of the International Conference on Telecommunications and Multimedia*, Chania, Greece, July 2012.

[9] Y. Zhu and M. Ammar, "Algorithms for assigning substrate network resources to virtual network components," in *Proceedings of the 25th IEEE International Conference on Computer Communications (INFOCOM '06)*, Barcelona, Spain, April 2006.

[10] N. Mosharaf, K. Chowdhury, M. R. Rahman, and R. Boutaba, "Virtual network embedding with coordinated node and link mapping," in *Proceedings of the 28th IEEE Conference on Computer Communications (INFOCOM '09)*, pp. 783–791, Rio de Janeiro, Brazil, April 2009.

[11] S. Srikantaiah, A. Kansal, and F. Zao, "Energy aware consolidation for cloud computing," in *Proceeding of the 2008 Conference on Power Aware Computing and System*, USENIX Association, 2008.

[12] B. Guenter, N. Jain, and C. Williams, "Managing cost, performance, and reliability tradeoffs for energy-aware server provisioning," in *Proceedings of the 26th IEEE Conference on Computer Communications (INFOCOM '11)*, pp. 1332–1340, Shanghai, China, April 2011.

[13] Q. Huang, F. Gao, R. Wang, and Z. Qi, "Power consumption of virtual machine live migration in clouds," in *Proceedings of the 3rd International Conference on Communications and Mobile Computing (CMC '11)*, pp. 122–125, Qingdao, China, April 2011.

[14] H. Liu, H. Jin, C. Xu, and X. Liao, "Performance and energy modeling for live migration of virtual machines," *Cluster Computing*, vol. 16, no. 2, pp. 249–264, 2013.

[15] S. Su, Z. Zhang, X. Cheng, Y. Wang, Y. Luo, and J. Wang, "Energy-aware virtual network embedding through consolidation," in *Proceedings of the IEEE INFOCOM Workshop on Communications and Control for Sustainable Energy Systems: Green Networking and Smart Grids (INFOCOM '12)*, Orlando, Fla, USA, April 2012.

[16] J. F. Botero, X. Hesselbach, M. Duelli, D. Schlosser, A. Fischer, and H. D. Meer, "Energy efficient virtual network embedding," *IEEE Communications Letters*, vol. 16, pp. 756–759, 2012.

[17] J. F. Botero and X. Hesselbach, "Greener networking in a network virtualization environment," *Computer Networks*, vol. 57, pp. 2021–2039, 2013.

[18] L. Chiaraviglio, D. Ciullo, M. Mellia, and M. Meo, "Modeling sleep mode gains in energy-aware networks," *Computer Networks*, vol. 57, pp. 3051–3066, 2013.

[19] A. Adelin, P. Owezarski, and T. Gayraud, "On the impact of monitoring router energy consumption for greening the Internet," in *Proceedings of the 11th IEEE/ACM International Conference on Grid Computing (Grid '10)*, pp. 298–304, Bruxelles, Belgium, October 2010.

[20] R. Bolla, F. Davoli, R. Bruschi, K. Christensen, F. Cucchietti, and S. Singh, "The potential impact of green technologies in next-generation wireline networks: is there room for energy saving optimization?" *IEEE Communications Magazine*, vol. 49, no. 8, pp. 80–86, 2011.

[21] C. Barnhart, N. Krishnan, and P. H. Vance, "Multicommodity flow problems," in *Encyclopedia of Optimization*, pp. 2354–2362, Springer, 2009.

[22] L. Chiaraviglio, M. Mellia, and F. Neri, "Reducing power consumption in backbone networks," in *Proceedings of the IEEE International Conference on Communications (ICC '09)*, Dresden, Germany, June 2009.

[23] J. Chabarek, J. Sommers, P. Barford, C. Estan, D. Tsiang, and S. Wright, "Power awareness in network design and routing," in *Proceedings of the 27th IEEE Communications Society Conference on Computer Communications (INFOCOM '08)*, pp. 457–465, April 2008.

[24] LXC Linux Containers, http://lxc.sourceforge.net.

[25] Quagga Routing Software, http://www.quagga.org.

[26] Linux Bridge, http://sourceforge.net/projects/bridgelinux/.

[27] LSA Generator, LSA Generator Software SPOOF, http://www.cs.ucsb.edu/rsg/Routing/download.html.

[28] RUDE and CRUDE, http://rude.sourceforge.net.

[29] Wireshark, http://www.wireshark.org.

[30] V. Eramo, M. Listanti, N. Caione, I. Russo, and G. Gasparro, "Optimization in the shortest path first computation for the routing software GNU Zebra," *IEICE Transactions on Communications B*, vol. E88, no. 6, pp. 2644–2649, 2005.

Hindawi Publishing Corporation
Journal of Electrical and Computer Engineering
Volume 2014, Article ID 190586, 14 pages
http://dx.doi.org/10.1155/2014/190586

Research Article

Energy-Aware Base Stations: The Effect of Planning, Management, and Femto Layers

G. Koutitas,[1,2] L. Chiaraviglio,[3] Delia Ciullo,[4] M. Meo,[5] and L. Tassiulas[1]

[1] *Computer Engineering and Telecommunications, University of Thessaly, 38221 Volos, Greece*
[2] *School of Science and Technology, International Hellenic University, Thermi, Greece*
[3] *DIET Department, University of Roma-La Sapienza, Via Eudossiana 18, Rome, Italy*
[4] *Mobile Communications Department, Eurecom, Sophia Antipolis, 450 Route des Chappes, 06410 Biot, France*
[5] *Telecommunication Networks Group, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy*

Correspondence should be addressed to G. Koutitas; george.koutitas@gmail.com

Received 21 October 2013; Accepted 3 January 2014; Published 30 March 2014

Academic Editor: Vincenzo Eramo

We compare the performance of three base station management schemes on three different network topologies. In addition, we explore the effect of offloading traffic to heterogeneous femtocell layer upon energy savings taking into account the increase of base station switch-off time intervals. Fairness between mobile operator and femtocell owners is maintained since current femtocell technologies present flat power consumption curves with respect to served traffic. We model two different user-to-femtocell association rules in order to capture realistic and maximum gains from the heterogeneous network. To provide accurate findings and a holistic overview of the techniques, we explore a real urban district where channel estimations and power control are modeled using deterministic algorithms. Finally, we explore energy efficiency metrics that capture savings in the mobile network operator, the required watts per user and watts per bitrate. It is found that the newly established pseudo distributed management scheme is the most preferable solution for practical implementations and together with the femotcell layer the network can handle dynamic load control that is regarded as the basic element of future demand response programs.

1. Introduction

The traffic proportional power consumption of information and communication technology (ICT) equipment is regarded as a promising technological growth for the improvement of the overall energy efficiency of the sector [1]. Despite the fact that individual components are already operating in different power states, according to the input load, the aggregated power profile of the network presents low correlation with traffic. This results in a power waste and a low efficiency, especially during off-peak hours. In a cellular network, base stations (BSs) are regarded as the weakest part in terms of energy efficiency since they comprise network critical physical infrastructure (NCPI) that introduces high *no load* losses (i.e., the losses that are associated with cooling and power units, which do not depend on the traffic) [2]. Recent 3 Generation Partnership Program (3GPP) regulations and new standards indicate the required steps for traffic proportional mobile network power consumption [3, 4] that incorporate BS management schemes and radio network planning phases [5, 6]. Depending on the network layout and the traffic pattern, it has been shown that significant savings can be achieved by reducing the number of active BSs by introducing sleep modes [7].

This paper aggregates and simulates the most important BS management schemes found in the literature on three different network topologies that were derived by an optimization algorithm. The BS management schemes are the centralized, the distributed, and the pseudo distributed schemes, each one presenting crucial differences in terms of savings and complexity. The network topologies were derived by the *minimum transmitter*, the *minimum power consumption*, and a *hybrid* planning strategy that are regarded as the most dominant planning strategies for future networks.

In addition, the paper explores the effect of heterogeneous networks (HetNets), such as femtocell, on energy savings and dynamic load control.

It is shown that the hybrid planning strategy coupled together with the pseudo distributed BS management scheme and also taking advantage of the HetNet can provide not only great OPEX reductions to the mobile network operator but also dynamic load control that is regarded as a fundamental element of future demand response programs. Basic capital expenditure (CAPEX) and operational expenses (OPEX) computations are performed.

The paper is organized as follows. In Section 2 a description of the related work is presented. Section 3 presents the network model used together with the power control scheme, the femto user association rules, and the propagation models. Section 4 presents the network architecture and the concept of critical and flexible BSs. Section 5 presents the BS management algorithms and Section 6 presents the used energy efficiency metrics. Finally, Section 7 explores the simulation results and Section 8 concludes our work.

2. Related Work on Energy Efficiency

A centralized BS management scheme for a macro-/microarchitecture is presented in [7]. In [8] the authors examine centralized management algorithms based on the least load and cell overlap criteria, applied upon a real network configuration of central London. The concept of energy partitions for Self-Optimized Network (SON) operation is given in [9, 10]. In [9] a pseudo distributed control scheme is introduced and applied upon a hexagonal network configuration. The partition in that case is assumed to comprise the flexible stations under the administrative domain of the critical ones. In [11, 12] authors investigate traffic proportional network power consumption for 3G networks utilizing optimization algorithms. In [13] centralized and distributed algorithms are used for a microarchitecture. A different approach to energy management is given in [14, 15] presenting the case of cooperation between two mobile operators. In [16] authors examine the BS planning effect showing that micro-BS architectures can increase the network's performance. Similar observations are also derived in [17]. In addition, the femtocell layer is shown to be of minor importance if one considers the network in large scale and does not impose any BS management schemes. In [18] the authors prove that a microcell based network provides the important gains, but the network does not present traffic proportional characteristics. In [19] the effect of service rate on the energy consumption of mobile network that supports switch-on/-off schemes is analyzed. Physical layer energy efficient techniques are presented in [20] focusing on physical layer and power control techniques. Finally, BS power consumption models that can be used for the estimation of the network's energy demand are investigated in [21–23]. A quantitative analysis of the effect of femtocell layer (HetNet) on a theoretical hexagonal network configuration is presented in [24, 25]. The authors examine CAPEX and OPEX characteristics according to femtocell deployment densities in a macrofemto architecture. In this paper we extend the investigation of the effect of three different planning and management procedures on a real network configuration and we explore the feasibility of a pseudo distributed management algorithm together with the effect of the femtocell layer on energy efficiency and savings. We perform simulations for two different user-femtocell association rules to capture realistic and theoretically maximum gains derived from the HetNet. Finally, we explore various energy efficiency metrics to capture not only the energy savings at the operator side but also the energy efficiency of offered service. The paper gives a more holistic view compared to the available literature in the area of network planning, BS management, and offloading for energy efficiency.

3. The Cellular Network

3.1. Description of Scenario. An area A of central London is under investigation and is presented in [18]. The area is a central urban district modeled by a vector map. The vector map describes the facets of the buildings that are incorporated in the ray tracing algorithm for channel estimation. It is a rectangular area of size 1.8 Km × 1.8 Km with equally spaced points of distance 50 m creating a set of possible coordinates CRD. We partition the coordinates in two different sets, that is, the outdoor CRD_0 and indoor coordinates CRD_I with $\mathrm{CRD}_0 \cap \mathrm{CRD}_\mathrm{I} = \varnothing$ and $\mathrm{CRD}_0 \cup \mathrm{CRD}_\mathrm{I} = \mathrm{CRD}$. Within A there is a set of sites S, with identifiers $s_j \in S$, $j \in \{1, 2, \ldots N_{\mathrm{Tx}}\}$ where N_{Tx} is the number of BSs. These sites are *macrocell* and *microcell* BSs that are under the administrative domain of the mobile operator. Furthermore, we assume that there are opportunistically deployed *femtocell access points* (FAPs) in the network within the users' premises. These stations form a set $f \in F$, $f \in \{1, 2, \ldots N_{\mathrm{FTx}}\}$ where N_{FTx} is the number of FAPs.

We define as $M_a \subset S$ and $M_i \subset S$ the set of macrocells and microcells, respectively. Each site, $j \in S$, has 5 characteristic values, representing the position, the type of station, the maximum transmit power level, and the antenna gain following the notation in [18]. In a similar approach, the FAPs, $f \in F$, are characterized by the position which is defined by Cartesian coordinates $\{x_f, y_f\} \in \mathrm{CRD}_\mathrm{I}$, the height H_f, the maximum transmit power level $P_f = 0.5$ W, and the antenna gain $G_f = 2$ dBi. Table 1 reports the main notation used in this work.

We model the users (described by a set U) and their properties as a population in the network topology using the user equivalent definition [8]. Each user is characterized by the location, the antenna gain, and the type of service requested indicating the required data rate and minimum signal to noise ratio threshold (Eb/Io $= \delta_i$). It is assumed that the user positions, during one simulation, are constant with time and uniformly distributed over CRD_o. Since the outdoor coordinates in the urban district are not uniform due to the building architecture, the uniform distribution of users over CRD_o yields more concentration in wide roads and

Table 1: Main notation.

Symbol	Explanation	Value
A	Area under investigation	1.8 ×1.8 Km
CRD	Set of possible coordinates	Equally spaced points of distance 50 m
S	Set of sites	Cartesian coordinates inside A
M_a	Set of macrocells	23 possible sites
M_i	Set of microcells	43 possible sites
N_{Tx}	Number of base stations (macrocells and microcells)	69
F	Set of femtocells access points	Cartesian coordinates inside A
N_{FTx}	Number of femtocells access points	25–75
$\{x_f, y_f\}$	Cartesian coordinates of femtocell f	Coordinates inside A
H_f	Height of femtocell f	1.5 m above floor
P_f	Maximum transmit power level for femtocell f	$\{0, 0.5\}$ W
P_j	Maximum transmit power level for base station j	$\{0, 10, 20, 30, 40, 50\}$ W (macro) $\{0, 1, 2, 3, 4, 5\}$ W (micro)
G_f	Antenna gain for femtocell f	2 dBi
U	Set of users	Cartesian coordinates inside A,with $\lvert U\rvert = 847$
δ_i	Signal to noise ratio threshold for user i	5 dB for voice, 2.5 dB for video, and 2 dB for web service
$G(U, S)$	Affinity graph for users and base stations	Set of edges between $U$and S
$G(U, F)$	Affinity graph for users and femtocells	Set of edges between U and F
B_j	Best server site	Identifier of the site
p_j^c	Power associated with the control channel of base station j	15% of the total power
HN_j	Soft handoff site	Identifier of the site
q	Soft handoff margin	5 dB
l_{ij}	Path loss from user i to base station j	Variable
CV_j	Coverage umbrella of base station j	Variable
x_{ij}	Variable indicating if user i is associated with base station j	$\{0, 1\}$
G_i	Antenna gain of base station i	10 dB (macro), 3 dB (micro)
G_i	Antenna gain of user i	2 dBi
ω	Orthogonality factor	0.6
ν_i	Voice activity	0.58
μ_i	Carrier to interference plus noise ratio for user i	Variable
π_p	Minimum sensitivity threshold	−115 dBm
n	Thermal noise	−174 dBm/Hz
$\overline{P}_j$	Mean transmit power of base station j	Variable
r_i	Required rate for user i	12 Kbps for voice, 64 Kbps for vide, 144 Kbps for web service
C	Chip rate	3.87×10^6 cps
W_j	Total consumed power for base station j	Equation (8a)
W_f^F	Total consumed power for femtocell f	10 W
R_j	Maximum data rate processed by base station j	14 Mbps (macro), 4 Mbps (micro)
a_j	Binary variable indicating if base station j is used	$\{0, 1\}$
$U^{\max}$	Set of users during peak hour	Cartesian coordinates inside A
$N^{\max}$	Number of users during peak hour	847
$N^{\min}$	Minimum number of users	40
t	Current time	Variable
t_ζ	Decision time	Variable
T	Maximum time	24 h

Table 1: Continued.

Symbol	Explanation	Value
Ω_k	Set of base stations that are controlled by a critical base station k in the pseudo distributed algorithm	Variable
ε	Load threshold	0.3 for macros and 0.03 for micros

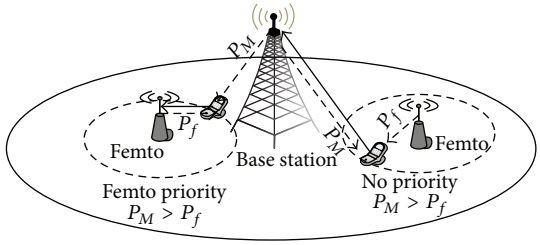


Figure 1: User association rules. Macro/micro architecture and femto access.

open parks, capturing real-life scenarios. A snapshot SN is described by mapping each mobile user to a location in the area and the appropriate user characteristics. A bigraph, the *affinity graph*, is used to model the affinity of user equivalents to sites, $G(U, S)$ or $G(U, F)$, which is undirected with vertex set $V\{S \cup U\}$ or $V\{F \cup U\}$ such that there is an edge between users $u \in U$ and $s \in S$ or $f \in F$ if user u is under the coverage umbrella of site s or FAP f. A subgraph of the affinity graph, the *association graph*, represents the real association of users with the site based on the server selection criteria.

3.2. User Association Rules. It is assumed that the FAPs are in open loop access and thus users can be associated with the femtos according to capacity and signal constraints. We distinguish two cases (see Figure 1). In the first case, BSs and FAPs are treated in a balanced manner and the user is connected to $j \in S$ or $f \in F$ according to the best server criteria. This user association rule is named as *no priority*. In the second approach, users are assumed to prefer femtocell connection and are connected to a femto if and only if the received signal from the FAP satisfies quality of service (QoS) criteria. This user association rule is named as *femto priority*.

3.2.1. No Priority Rule. We define as the best server site B_j the site $j \in S$ or $j \in F$ that associates users $i \in U$ with B_j under the following criteria [26] ($l_{ij} < 1$ represents the path loss of user i from site j):

$$B_{j \in S \cup F} := \{i\} \longrightarrow \arg\max_i \left(G_j \cdot g_i \cdot p_j^c \cdot l_{ji}\right). \tag{1}$$

In the above formulation $p_j^c = z \cdot P_j$ is the power associated with control channels and is a function of factor z (usually it is $z = 0.15$ describing 15% of radio frequency (RF) out power associated with control channels [26]). To model soft handoff mechanism, we define as the soft handoff site HN_j the site that associates $i \in U$ with HN_j, under the following criteria (q is the soft handoff margin):

$$\mathrm{HN}_{j \in S \cup F} := \left\{u_i \in U : \exists k \neq j \text{ such that } \left|l_{ij} G_j p_j^p - l_{ik} G_k p_k^p\right| \leq q,\ q = 5\ \mathrm{dB}\right\}. \tag{2}$$

Thus, in the association graph, each vertex can have a degree of 1 or larger. By denoting $x_{ij} = 1$ if user i is connected to site j and 0 otherwise, we impose the following constraints:

$$\sum_{j \in S \cup F} x_{ij} = 1, \quad \forall i \in B_j,$$
$$\sum_{j \in S \cup F} x_{ij} > 1, \quad \forall i \in \mathrm{HN}_j. \tag{3}$$

The user is assumed to be within the coverage umbrella (set of users belonging to CV_j) of a specific cell if and only if QoS criteria are satisfied. In a mathematical form it is

$$\mathrm{CV}_j := \left\{ \begin{array}{c} u_i \in B_j, \mathrm{HN}_j : l_{ij} G_j p_j^c \geq \pi_p,\ l_{ji}^{\uparrow} G_i p_i^{\uparrow} \geq \pi_p^{\uparrow}, \\ \gamma_{ij} = \dfrac{l_{ij} G_j p_{ij}}{n + \omega l_{ij} G_j \left(\overline{p}_j - \nu_i p_{ij}\right) + \sum_{k \in U \setminus \mathrm{HN}_j, k \neq j} l_{ik} G_k \overline{p}_k} \geq \mu_i \end{array} \right\}, \tag{4}$$

where ω is the orthogonality factor (~0.6), n is the thermal noise, ν_i is the voice activity (~0.58), μ_i is the required carrier to interference plus noise ratio, π_p is a minimum threshold [26], and $\overline{p}_j$ is the mean transmit power for base station j. Indicator $\uparrow$ symbolizes the uplink channel.

3.2.2. Femto Priority Rule. This rule gives priority to the FAPs. It is used to model the upper bound of energy savings that can be obtained by femtocells. Under this rule the users in the network are categorized in two sets. In the first set, $U^F \subset U$, the users are connected to FAPs if they receive adequate power from the FAPs; that is, $l_{if} G_f p_f^c \geq \pi_p$, $f \in F$. To simplify computations, we assume that the FAPs and the macro-/micronetwork operate in different channels and thus interference between the multitier networks is negligible [27]. The remaining users are treated similar to the equations in (1), (2), and (4) taking into account only BSs $j \in S$.

3.3. Channel Estimation Algorithms. Due to the complex nature of the urban scenario, two different channel estimation algorithms were considered. A deterministic 3D ray tracing

algorithm was used for field predictions between a user and a macro-/microstation and an empirical propagation model was used to model the channel gain between a FAP and a user. The empirical model was preferred since the indoor clutter within the houses is opportunistic and thus it cannot be deterministically modeled.

3.3.1. Ray Tracing Algorithm. The used ray tracing algorithm is presented in [18]. It is based on an intelligent preprocessing of the database algorithm where the vector data of the buildings is described by tiles and segments. A ray is propagated in the environment by multiple reflections, diffractions, or a combination of the above mechanisms between the tiles and segments, until it reaches the receiving point.

3.3.2. Empirical Algorithm for Femtos. To model the indoor to outdoor propagation, the Keenan-Motley formulation was used as presented in [28, 29]. The propagation loss between a user and a FAP is

$$l^{\mathrm{dB}} = 20\log_{10}\left(\frac{4\pi f}{c}\right) + 20\log_{10} d + q_{\mathrm{in}} W_{\mathrm{in}} + \\ + q_{\mathrm{ex}} W_{\mathrm{ex}} + F n^{((n+2)/(n+1)-0.46)}. \tag{5}$$

All parameters are explicitly defined in [28]. For the purpose of the simulation results it was assumed that the random variables are uniformly distributed with $q_{\mathrm{in}} \leftarrow U_n[1, 10]$, $q_{\mathrm{ex}} \leftarrow U_n[1, 4]$, $n \leftarrow U_n[1, 5]$, $W_{\mathrm{in}} = 5\,\mathrm{dB}$, $W_{\mathrm{ex}} = 7\,\mathrm{dB}$, $F = 18\,\mathrm{dB}$, and $f = 2\,\mathrm{GHz}$.

3.4. Power Control Mechanism. The associated downlink power from BS with the served users is computed according to a perfect signal to interference noise ratio (SINR) based power control scheme. Given the required SINR (δ_i) for each user, the code division multiple access (CDMA) spreading factor, the bit rate r_i, the chip rate C of the system (3.84 Mcps), and the required transmit power of station j to satisfy QoS of user i are given by [18]

$$p_{ij} = \delta_i \nu \frac{r_i}{C}\left[(1-\omega) P_j l_{ij} G_j + \sum_{k \in U \setminus \mathrm{HN}_j, k \neq j} l_{ik} G_k \overline{p}_k + n\right] \frac{1}{l_{ij} G_j}, \tag{6}$$

where l is the path loss, G represents the antenna gain, n is the thermal noise, and P is the transmit power. This process works in an iterative way taking into account power allocations in other cells explicitly. The iterative process is continued until convergence is achieved. The transmit power converges to the minimum possible value for all BSs [18]. The mean transmit power of the base station is given as the summation of the control and the traffic channels:

$$\overline{p}_j = \sum_{i \in \mathrm{CV}_j} \nu_i p_{ij} + p_j^c \leq P_j. \tag{7}$$

3.5. Base Station Characteristics

3.5.1. Power Consumption of Base Stations. An empirical model [22, 23, 30] describing a traffic proportional power consumption characteristic of BS was used:

$$W_j = b \cdot \overline{p}_j + c, \quad j \in S, \tag{8a}$$

$$\begin{aligned} &\mathrm{Macro} :\longrightarrow b_{\mathrm{Ma}} = 22.6, \qquad c_{\mathrm{Ma}} = 412.4\,\mathrm{W}, \\ &\mathrm{Micro} :\longrightarrow b_{\mathrm{Mi}} = 5.5, \qquad c_{\mathrm{Mi}} = 32\,\mathrm{W}. \end{aligned} \tag{8b}$$

The FAPs are residential stations with low power consumption characteristics which are independent of the number of users served. For the purpose of our investigation, it was assumed that [31]

$$W_f^F = 10\,\mathrm{W}, \quad f \in F. \tag{9}$$

3.5.2. Capacity of Base Stations. The number of users that can be served by a BS depends on the available resources. These are mainly related to transmit power limitations and data rates that can be processed by the IT equipment. The used capacity constraints are

$$\mathrm{Load}_j = \frac{\overline{p}_j}{P_j} \leq 1, \quad j \in S, \tag{10a}$$

$$\sum_{i \in \mathrm{CV}_j} x_{ij} r_i \leq R_j, \quad j \in S, \tag{10b}$$

$$\sum_{i \in \mathrm{CV}_j} x_{ij} \leq 4, \quad j \in f. \tag{10c}$$

In the above equation, $R_j = 14\,\mathrm{Mbps}$, $j \in M_a$, $R_j = 4\,\mathrm{Mbps}$, $j \in M_i$, and r_i is the bit rate of user i. For the FAPs it was assumed that they can serve 4 users simultaneously [27].

3.6. Traffic Profile. The examined traffic profile is presented in Figure 2 [32]. During low traffic hours, met at around 3 am, the normalized traffic is approximately 15% of the maximum. From the traffic curve it can be observed that, on average, the network is approximately 57% occupied and that most of the time the network operates below 57%. This means that underutilization of the available resources is met. This indicates that serious underutilization occurs and energy-aware BS management schemes need to be put into place to save energy.

4. Network Architectures

According to the operator's objectives in terms of CAPEX and OPEX, different network planning strategies can be followed. For the purpose of our investigation three different network topologies were considered. Since the network planning is proven to be nondeterministic hard (NP-hard), the network topologies were derived by a genetic algorithm (GA) optimization technique. In particular, we consider the minimization of the number of BSs to provide a predefined QoS over the area (named as TX strategy), the minimization

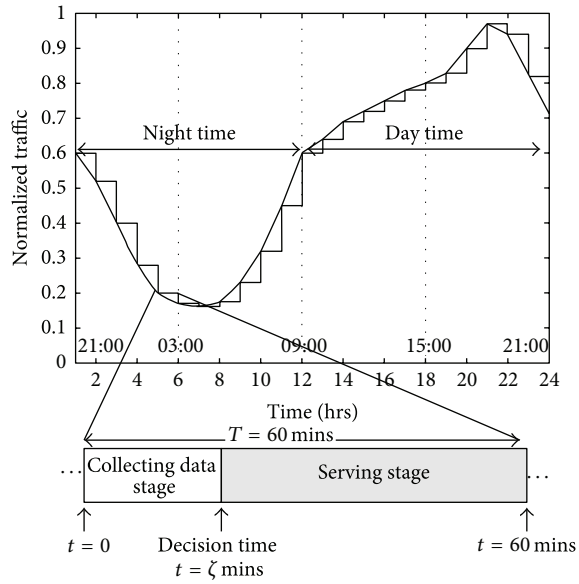


Figure 2: Traffic profile and pseudo real time management procedure.

Table 2: Set of deployed and managed base stations.

Planning	Set	Number of macros	Number of micros
TX	High traffic	5	1
	Low traffic	2	—
	Flexible stations	3	1
MP	High traffic	1	14
	Low traffic	1	7
	Flexible stations	—	7
HY	High traffic	4	6
	Low traffic	2	3
	Flexible stations	2	3

Table 3: Yearly saving (%) under different strategies.

	TX	MP	HY
No femto	12.6	8.8	22.4
25 no femto priority	15.0	11.0	25.9
25 femto priority	25.2	14.0	31.0
75 no femto priority	19.4	15.4	32.1
75 femto priority	43.6	22.0	42.0

of power consumption of the network for the same QoS (named as MP strategy), and finally a hybrid approximation (named as HY strategy) as shown in [18]. The GA was used to search the decision variables of the problem that are the base station positions, a_j (a binary indicator indicating if a BS exists at a given location), the type of the station, T_j (binary indicator indicating macro- or micro-BS), and the transmit RF power, P_j [18]. The radio planning procedure was implemented for two traffic scenarios. The first planning is performed considering the peak traffic periods while the second one is applied to low traffic periods. The final network topology comprises a set of BSs $D \subset S$.

Three sets of BSs are distinguished in the network. Set MD $\subset D$ is the *critical* base stations that cannot be set to sleep mode, but it participates in the on-/off-management game by providing load information and by coordinating a subset of BSs. Set FD $\subset D$ of BSs can change its state of operation and is named as flexible stations, and finally set $\Omega_k \subset$ FD, k is the index, is the set of BSs that is controlled in the pseudo distributed algorithm by a critical BS, $k \in$ MD. The set Ω_k is defined according to the cell overlap parameters of the cells. By indicating Ψ_{kj} with $0 \leq \Psi_{kj} \leq 1$, the normalized cell overlap value is between stations $k \in$ MD and $j \in$ FD and then set Ω_k is defined as

$$\Omega_{k \in \text{MD}} = \{j\} \longrightarrow \arg\max_{j \in \text{FD}} \left(\Psi_{kj}\right). \tag{11}$$

The aim of the two planning phases is to define the set of flexible BSs for each network configuration. It must be noted that FAPs are not included within the radio planning procedure since they are not managed by the operator. The examined network configurations are shown in Table 1.

4.1. High Traffic Planning. Following the definitions presented in Section 3.1, let $U^{\max}$ be the set of users during peak hour, with cardinality $N^{\max}$. The high traffic planning is performed to fulfill coverage and capacity issues for the maximum possible number of users.

4.1.1. Minimum Transmitter (TX). The TX strategy minimizes the number of BSs and the objective function can be written as

$$\text{TX} \longrightarrow \min \sum_{j \in S} a_j. \tag{12}$$

The final network configuration comprises BSs $j \in D^{\text{TX}}$. This network configuration yields to a network with a small number of high power BSs (mainly macros; Table 2).

4.1.2. Minimum Power Consumption (MP). This strategy corresponds to a network configuration with the minimum power demands. The objective is

$$\text{MP} \longrightarrow \min \sum_{j \in S} W_j. \tag{13}$$

The decision variables and constraints are the same with the TX strategy. The final network configuration is described by $j \in D^{\text{MP}}$. This network configuration yields to a network comprising a large number of low power base stations (mainly micros; Table 2).

4.1.3. Hybrid Scenario (HY). *Hybrid network* is the topology that falls in between the TX and MP. Thus there are a smaller number of BSs compared to TX and higher power consumption compared to MP. The HY network configuration is a balanced macro-/micronetwork that is closer to real network deployments. The final network configuration is described by $j \in D^{\text{HY}}$ (see Table 3). A more detailed description of the HY strategy is given in [18].

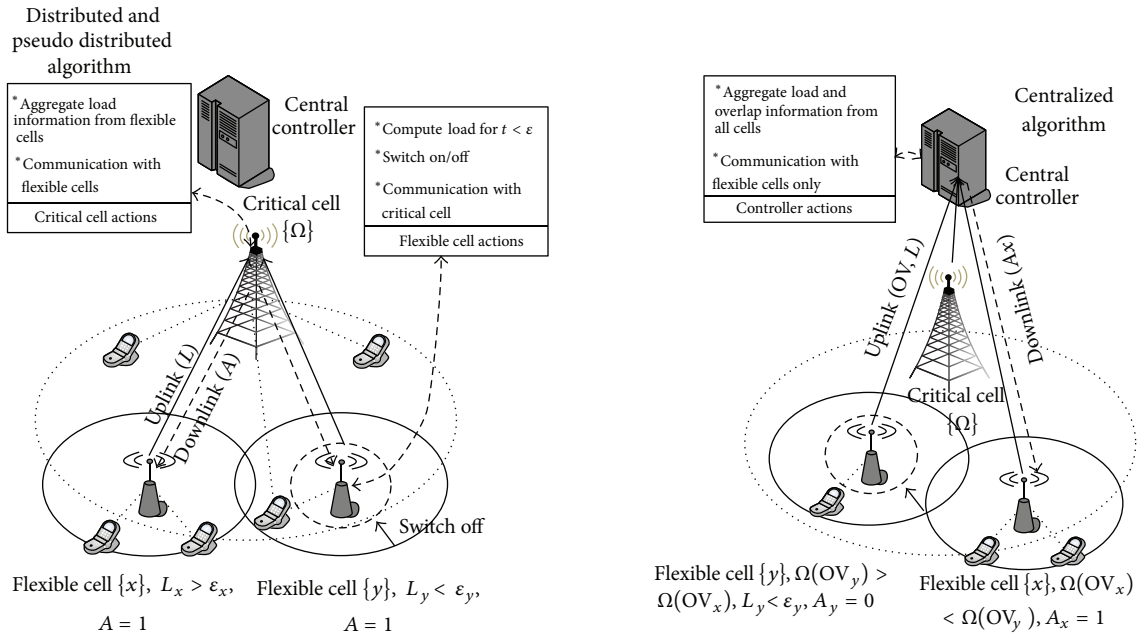


Figure 3: Base station management architectures.

4.2. Low Traffic Planning. Most of the BSs at low traffic periods are underutilized and the operation of NCPI equipment reduces the energy efficiency of the network and increases the energy waste. To define the set of flexible BSs, we perform a minimum transmitter optimization strategy for each configuration D^X. It is similar to the minimum transmitter planning strategy, but the optimization algorithm searches the state space of D^X instead of the initial population of possible base station locations (S) and provides the topology $\mathrm{MD}^X \subset D^X$ that satisfies QoS and coverage for the minimum number of users $U^{\min}$. The objective function for that case is

$$\mathrm{MD}^X \longrightarrow \min \sum_{j \in D^X} a_j. \tag{14}$$

All constraints remain the same. For the off-peak planning phase, the optimization algorithm searches only the a_j variables whereas the type of BSs T_j and the maximum transmit power P_j remain the same as in D^X.

The set of flexible BSs, FD^X, is described as $\mathrm{FD}^X = D^X \setminus \mathrm{MD}^X$. These stations model the degrees of freedom of the mobile network operator during the on-/off-management scheme.

5. Base Station Management Schemes

Base station management schemes can be performed in a centralized, distributed, or pseudo distributed manner. The control schemes present differences in terms of performance and complexity and are implemented by the Operation, Administration, and Management (OAM) part of the network [3, 8, 13, 14].

5.1. Centralized Management. This scheme is performed by the central controller of the network and decision is taken from OAM (Figure 3, right plot). The controller falls within the administrative domain of the mobile operator and captures critical data of the network. Two centralized approximations are investigated in this paper and these are the *least load* (LL) and the *most overlap* (OV) algorithms [8]. All heuristics start by considering a topology in which all BSs are powered on. Then the algorithm checks iteratively if a given BS can be turned off. At each iteration, the considered BS is removed from the topology. User coverage and BS capacity are then recomputed on the residual topology. If both coverage and capacity are still fulfilled, then the selected BS is definitively powered off. Algorithm 1 reports a schematic description of the heuristics. The complexity of the centralized scheme is equal to $O(\overline{\overline{D}})$ where D is the set of BSs of the network.

The least load (LL) strategy is based on the load sustained by the BSs in the considered deployment. Specifically, the BSs are selected starting from the least loaded one. The rationale is that low loaded BSs are more likely to be switched off first, avoiding frequent off-/on- transitions. During the LL algorithm the flexible BSs report to the controller their load and the controller decides if they can be switched off. A binary command $A = \{0, 1\}$ is used to describe if a BS should switch off ($A = 0$) or on ($A = 1$) according to coverage and capacity issues.

The most overlap (OV) strategy takes instead into account the overlapping coverage areas existing among neighboring BSs. The intuition is that, in dense deployments, several BSs are necessary to provide capacity during the peak hours, but they are redundant during low traffic periods. In this case,

```
(1) sort BS(BS array, order type);
(2) for j = 1; j ≤ size(BS array); j + + do
(3)     if BS array[j].id ∈ FD^x then
(4)         disable BS(BS array[j]);
(5)         user coverage = compute coverage(BS array);
(6)         BS capacity = compute capacity(BS array);
(7)         if (check coverage(user coverage) == false) ‖
            (check capacity(BS capacity) == false) then
(8)             enable BS(BS array[j]);
(9)         end if
(10)    end if
(11) end for
```

Algorithm 1: Centralized management schemes.

```
Symbols — **Distributed ^^Pseudo distributed
(1) while t < t_ζ, compute L_j, ∀j ∈ B
(2) if t = t_ζ, compute L̄_j ~ ∑ L_j/# samples
(3) **if L̄_j ≤ ε_j, j ∈ FD set P_j = 0 and if L̄_j > ε_j, j ∈ FD set P_j = 0
(4) ^^if L̄_j > ε_j, j ∈ MD set P_k = P_on, k ∈ Ω_j if L̄_k > ε_k or P_k = 0 if L̄_k ≤ ε_k
(broadcast command A = 1 to all flexible stations k ∈ Ω_j)
(5) ^^set P_j = P_on, ∀j ∈ MD and if L̄_j ≤ ε_j, j ∈ MD set P_k = 0, ∀k ∈ Ω_j
(broadcast command A = 0 to all flexible stations k ∈ Ω_j)
(6) continue until t = T
```

Algorithm 2: Distributed and pseudo distributed management schemes.

the BSs are sorted according to decreasing overlapping. For each BS, the overlapping is computed as the number of active users that can hear the current BS and at least another BS over the total number of users that can hear the current BS. We say that a user can hear a BS if the SINR threshold requirement on the control channel is satisfied.

5.2. Pseudo Distributed Management. The pseudo distributed algorithm authorizes critical BSs to initiate on-/off-command flow to the flexible stations that fall within their administrative domain. Switch-on/-off commands are given hierarchically by stations $k \in$ MD to flexible stations $j \in \Omega_k$ to maintain a smooth QoS during transitions. In this approach, the network is divided into distinct administrative clusters and communication overheads are only limited within each cluster (Figure 3, left plot). Each flexible station reports in the uplink the instantaneous load. The critical station aggregates data and it computes its own load. Decision parameter A (binary parameter $A = \{0, 1\}$) is broadcasted to all flexible stations $j \in \Omega_k$. $A = 1$ means that the flexible stations are allowed to change their state, whereas $A = 0$ means that all flexible stations in the administrative domain of the critical station should switch off. The pseudo code of this scheme is given in Algorithm 2. The complexity of the algorithm within each cluster k is $O(\overline{\overline{\Omega_k}})$.

5.3. Distributed Management. In the distributed approach, the flexible stations $j \in$ FD decide locally when to change their state of operation neglecting the traffic conditions in other cells of the network. In this approach, communication overhead between the base stations is minimized. The pseudo code of this scheme is given in Algorithm 2. The complexity for each BS is equal to $O(1)$. The algorithm is similar to the pseudo distributed management of Figure 3, but the decision of the state of each flexible BS is independent of the downlink parameter A of the critical stations. Thus, uplink and downlink communication between critical and flexible stations are minimized.

5.4. Algorithm Implementation. The BS management schemes adapt the network according to instantaneous load conditions. The flexible stations can change their state of operation based on pseudo real time network conditions. Pseudo real time conditions indicate that decisions are made at every predefined period of time, T, which can be defined by the mobile operator. To help the flexible stations to decide in which mode they will operate, we set a load threshold ε_j, $\forall j \in$ FD.

The concept of the proposed management algorithms is presented in Figure 2. A typical day, DT, is quantized into X periods with $X = \mathrm{DT}/T$. Each period X contains two stages: the collecting data stage for $0 < t < t_\zeta$ and the serving stage for $t_\zeta < t < T$. Within the collecting data stage flexible stations serve the users or they can only gather samples of traffic load to calculate the mean traffic load. The BSs might change their mode at a specific point in time ($t_\zeta \sim 5$ mins).

During the serving stage, the remaining active BSs operate to serve the users. The management algorithms guarantee a minimum outage probability p_0 for the active users $M(t)$.

5.5. Correlation to Recent Standards. According to 3GPP, the BS can operate into three distinct states to support the switch-on/-off scheme [3]. The no-ES (ES refers to energy saving) state is during high traffic where the BS cannot be set in sleep mode. The ESaving state is the switch-off state of the BS. The EScompensate state is the state where the BS is on to support coverage of nearby switched off BSs. Based on our categorization, critical stations are always in no-ES state or equivalently in EScompensate and flexible stations during low traffic periods can be at ESaving. 3GPP presents three control schemes. Centralized and distributed schemes are similar to the schemes described above. For the centralized case the OAM of the network initiates energy saving operation and determines the BS to be set in sleep mode. In the distributed case the OAM initiates energy saving operation of the network, but each BS decides on the state of operation (on/off) independently. Finally, a hybrid scheme is also described in [3] and it presents crucial differences to the proposed pseudo distributed case. In the 3GPP hybrid management, the BS switch-on/-off is implemented in collaboration with the OAM and the individual BSs targeting global optimization. The examined pseudo distributed case is related to a clustering solution that defines smaller administrative domains for BS management.

6. Modeling Energy Efficiency

6.1. Mobile Network Operator Power Consumption. This metric computes the network power consumption related to the administrative domain of the mobile operator (macro- and micro-BSs). The network power consumption is computed according to the power needs of the macro- and micro-BSs in (8a) and (8b):

$$\mathrm{MPC}\,[W] = \sum_{j\in D} W_j = \sum_{j\in S} a_j W_j. \tag{15}$$

The energy consumption over a given time window Γ is thus computed according to

$$\mathrm{MEC}\,[\mathrm{Wh}] = \sum_{j\in S} \int_0^{\Gamma} a_j W_j(t)\, dt. \tag{16}$$

6.2. Energy per User. This metric associates the required energy to serve a user in the network. It takes into account the MPC and also the power consumption of the femto stations. The metric is described as

$$\mathrm{WU}\,[W/\mathrm{user}] = \frac{\sum_{j\in S}\left[a_j W_j/\overline{\overline{B_j}}\right] + \sum_{f\in F}\left[W_f^F/\overline{\overline{B_f}}\right]}{\sum_{j\in S} a_j + \overline{\overline{F}}}. \tag{17}$$

The metric as presented in (17) is a heuristic and models the average watts per user per station in the network. The first part of the numerator models the watts per user at the administration domain of the mobile network operator (macrocells and microcells), the second part of the numerator gives the watts per user at the administration domain of the femtocell owners, and the denominator creates an average value per station for the whole network.

7. Simulation Results

7.1. Femtocell Layer Effect. This section investigates the effect of femtocells on the network performance. It is assumed that the network has 50 FAPs that are randomly placed in the area. The effect of femtocells is examined for the two user-to-femto association rules as presented in Section 3. In the first subplot of Figure 4 the number of users that are associated with the femtocell network is given for the two association rules. It is observed that for the femto priority rule the number of users that are associated with FAP is the same for the three planning strategies and this is expected. For the no-priority case it is observed that for the MP planning strategy a higher number of users are assigned to the FAPs, compared to the other configurations. This is because the MP network configuration comprises a large number of low power microstations that create almost identical channel characteristics with the FAPs. In the lower subplots of Figure 4 the comparison of the power consumption of the network for the no-femto and 50 femto cases is presented. It is observed that the femtolayer can reduce power needs of the network and the effect is more significant for the TX strategy. This is expected since the TX network configuration comprises a small number of high power stations (mainly macros) whose power consumption is proportional to the input load.

The femtocell effect on the energy efficiency metrics of (17) is presented in Figure 5. The figure presents comparison with the case of no femtos used. It is observed that the femto priority degrades the performance of the network for low traffic periods whereas it increases the efficiency during peak hours. On the other hand, the no-priority rule presents opposite characteristics. The efficiency is reduced since the network of the mobile operator serves less number of users but consumes a lot of power due to the no-load losses of the BSs. The effect of the no-load losses upon the energy efficiency is clear for the MP strategy (microstations) where the efficiency does not present significant changes due to the small losses of the microstations. Taking into account that the mean number of active users in the network is approximately 400, it can be concluded that, for the specific network, the femto priority user association rule is expected to increase the energy efficiency.

7.2. Base Station Management Effect. Figure 6 presents the reduction of the network power needs with respect to the case where no management is imposed, assuming no femtocells. It is observed that the centralized OV strategy yields the highest gains which are almost identical to the centralized LL strategy. The pseudo distributed management scheme is the scheme that presents lower gains compared to the centralized schemes but higher gains compared to the distributed one. This is expected since the critical BSs of

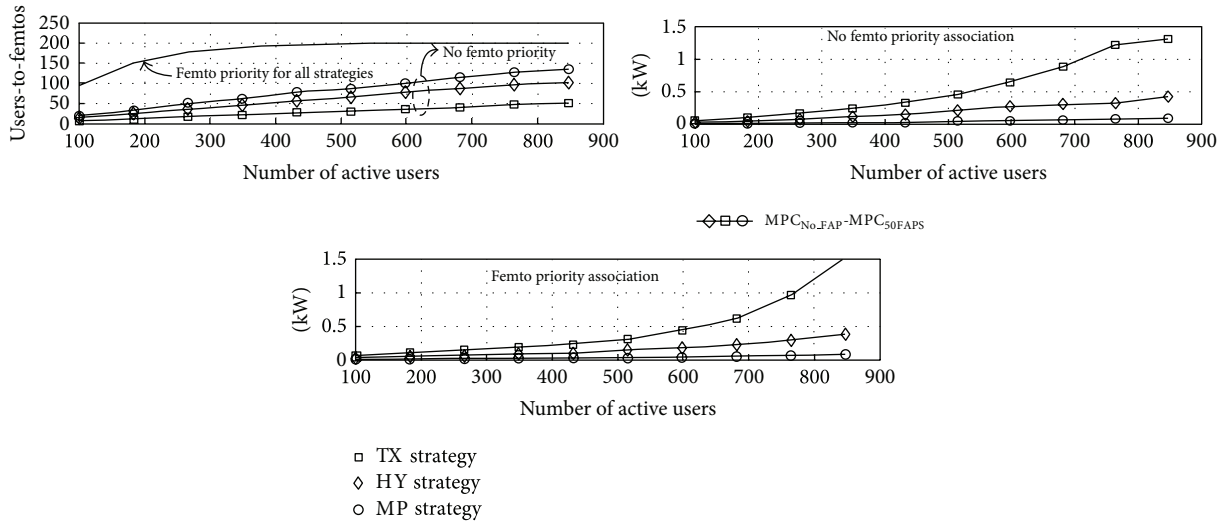


Figure 4: The upper subplot presents a number of users that are connected to FAPs for the two association rules. The lower subplots present the network power consumption of the network compared to the no femto case. 50 FAPs randomly positioned in the network were assumed.

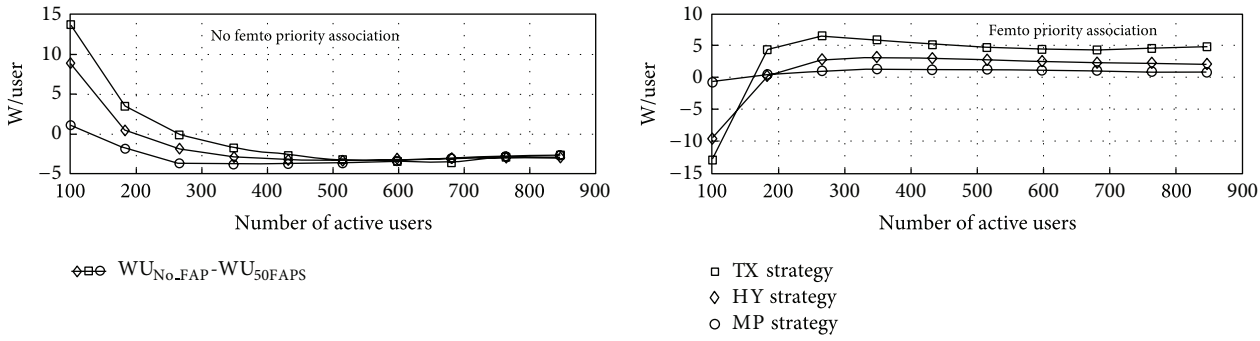


Figure 5: Energy efficiency metrics for the femto priority and the no priority cases compared to the no femto scenario. 50 FAPs were assumed.

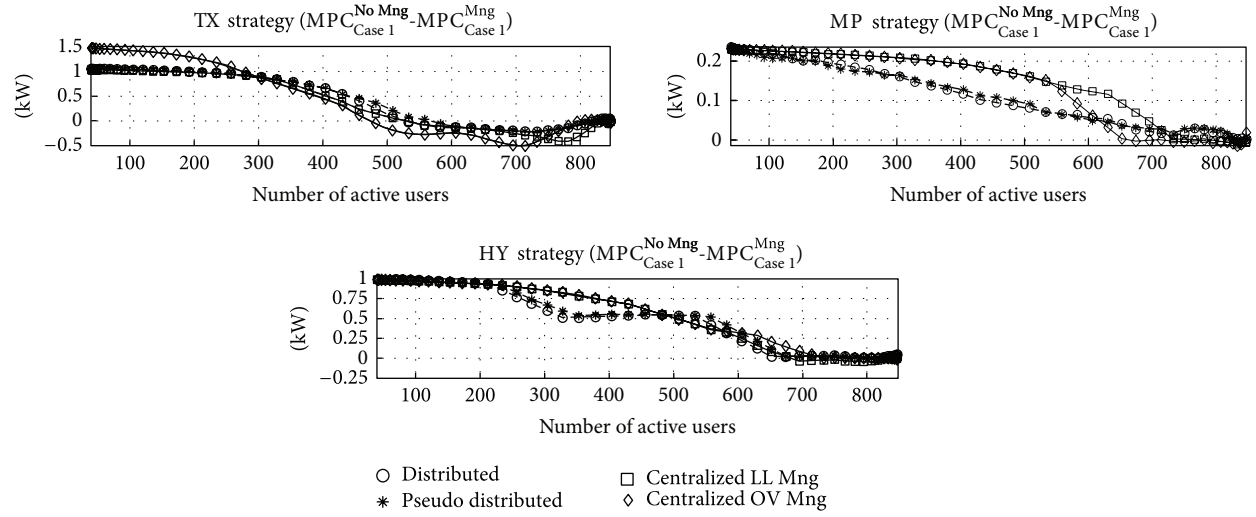


Figure 6: Comparison of network power consumption relative to the no management case.

the network coordinate more efficiently the subset of flexible BSs. It can be observed that for low traffic periods the savings are significant and this is because a large number of BSs are set to sleep mode. On the other hand, during peak traffic the management schemes converge to the performance of the no-management case. During each network transition, the coverage and the outage probability was almost equal to the minimum bounds imposed by the planning phase. The negative values observed for the centralized management indicate that if the network operates with small number of BSs at maximum load can yield worst energy performance compared to no management. This is obvious for 500 and 700 active users of the TX strategy for the centralized OV control. In that case, the network operates with some BSs at sleep mode, meaning that the active BSs are operating near maximum capacity and thus the total consumption is to be higher compared to the case where all BSs are on. Of course, this finding is sensitive to parameters b and c of (8a) and (8b). For the simulation results it was assumed that the load threshold for macrocell stations is equal to $\varepsilon = 0.3$ and $\varepsilon = 0.03$ for microcell stations [9]. Regarding the outage and network coverage, the centralized algorithm pushed the values to the lowest limits in order to keep the energy efficiency to the maximum value. The distributed and the pseudo distributed algorithms present a better outage and network coverage and a smoother transition between the different states. From the simulation results and by taking into account the increased overhead of the centralized management schemes, the pseudo distributed algorithm was assumed to better reflect real mobile network implementations.

7.3. Base Station Management and Femtocells. This section explores the effect of the femtocell layer on the BS management scheme. The pseudo distributed case is only examined since it was found that it is more appropriate for practical implementations. The scope of the analysis is to identify energy savings and network peak power reduction by increasing the number of femtos in the network. In addition, we investigate the increase of time in which the femtocell layer enables switch off management of BSs in the network as well as the effect on the coverage of the network.

The first simulation result explores the HY planning strategy over the daily traffic profile of Figure 2. The network power consumption and the number of active BSs for the case of 25 and 75 FAPs are plotted in Figure 7. By increasing the number of FAPs in the network, the BSs management scheme enables more flexible stations to be in sleep mode for larger periods of time with respect to the no femto case. The femtocell layer together with BS management scheme makes the power consumption of the network more proportional to traffic. The femto priority rule guarantees high energy savings, peak power reduction, and larger time periods in which the BSs are set in sleep mode. This is expected since the femtolayer absorbs a higher amount of traffic. The drawback of implementing the femto priority rule is that it is difficult to arrange special agreements between users and mobile operator. In addition, handover mechanisms for moving users are important issues.

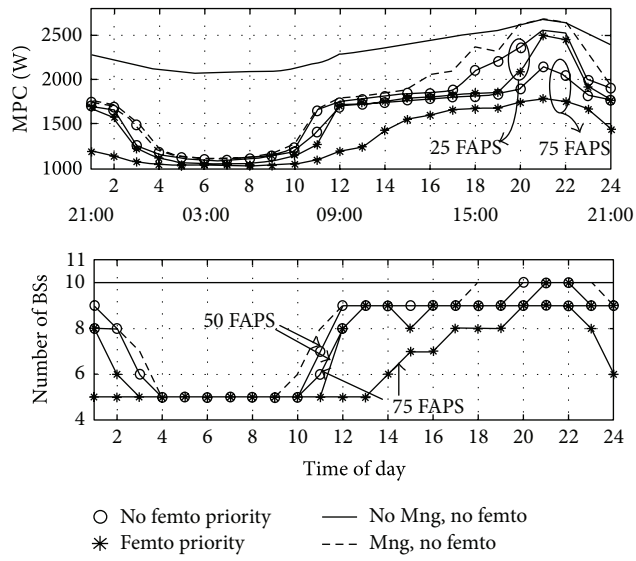


Figure 7: Energy consumption and number of active BSs over a typical day for the HY network configuration. Femtolayer effect is presented.

Figure 8 presents the energy savings and the peak power reduction as a function of active FAPs in the network. It can be observed that by increasing the number of FAPs the energy savings (16) are increased in an almost linear function. In addition, the peak power is reduced and this can have a significant effect if one considers the case where the BSs of the network are served by RES with limited power supply, in an island mode (net zero operation). By increasing the number of FAPs, it is also observed that the time interval during a day where BSs are set in sleep mode is significantly increased. This is more obvious for the MP strategy and the femto priority rule where almost the whole day the femtolayer enables more BSs to be set in sleep mode. Finally, it can be concluded that the coverage is improved by increasing the number of femtos. One can investigate similar to the smart grid demand response algorithms to enable the network to adapt its own power curve according to the available RES capacity. It is believed that the femtocell layer will open new business models for cellular networks that are supplied by RES.

7.4. Long-Term Operator Trajectories and Savings. In this section, we first briefly analyze the typical costs that operators face to deploy and maintain a network of BSs; then, we compute the potential savings achievable with the proposed sleep mode schemes. Specifically, we distinguish between operational expenditures (OPEX) and capital expenditures (CAPEX) that vary significantly between macro- and microbase station sites. Indeed, the cost of a macro-BS can be about seven to ten times higher than the one of a micro-BS [33].

In our cost assessment analysis, we use the typical costs for macro- and micro-BSs reported in Tables 1 and 2 of [33]. The CAPEX includes the cost to buy equipment (antennas,

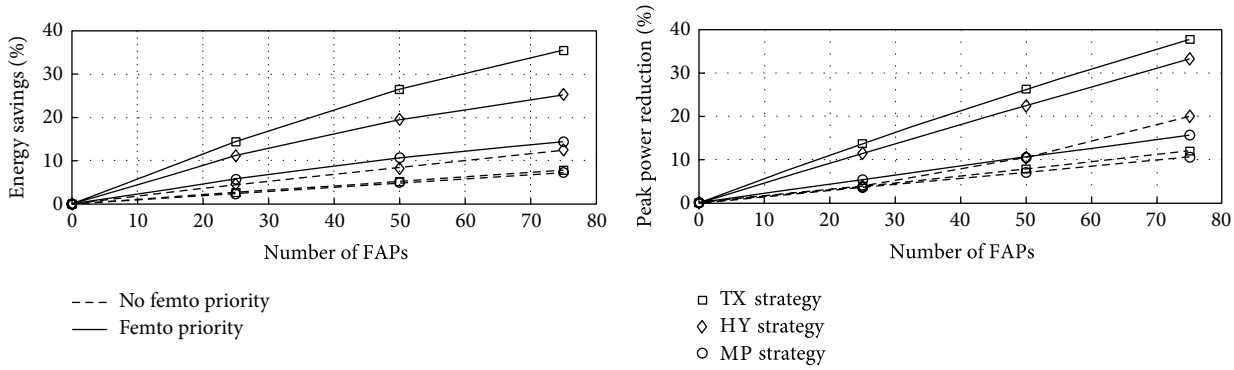


Figure 8: Energy savings, peak power reduction, and number of hours that BSs can be switched off and coverage relative to the no femtocell case. Comparisons concern both femto and no femto priority user association rule.

batteries, air conditioning, etc.) and the costs for site acquisition and deployment. The OPEX includes the costs for BS site rent, maintenance, power, and operation. Concerning the electricity cost, we assume that the cost of a KWh is equal to 0.1 Euro [30, 34]. Moreover, we can also compute the yearly saving in terms of tCO_2 emissions, by imposing that about 500 $grCO_2$ are emitted per 1 KWh [34].

In our analysis, first, we indicate which deployment among the ones obtained with the three strategies (TX, MP, and HY) is the best in terms of costs. Then, given the deployment, we compute the total yearly cost of the network for the following cases: (i) all the BSs are always on, and (ii) sleep modes are used. Finally we evaluate the energy that can be saved thanks to BSs' sleep modes.

In the following we show some numerical results in terms of savings obtained considering the three strategies under different network management schemes. First of all, we compute the CAPEX costs of the deployments. The most expensive deployment is the TX one (about 551 K€) since it includes more macro-BSs, while the MP costs about 262 K€ and the HY one costs about 498 K€. Thus, the more convenient strategy in terms of CAPEX is the MP one, since it comprises several but small BSs. The main drawback comes from the software perspective since the coordination of a large number of small cells is a complex procedure.

Then, we compute the savings obtained by adopting sleep mode schemes at the BSs. We report the percentage of achievable savings in Table 3, considering the cases in which (i) no femtocells are exploited in the network management scheme; (ii) 25 femtocells are used with/without any priority scheme; (iii) 75 femtocells are used with/without any priority scheme.

Observe that significant savings can be achieved with the TX and HY strategies, while savings are lower for the MP strategy. Moreover, it is interesting to note that savings are higher when the femto priority strategy is used and increase with the number of femtocells exploited by the network.

We also compute the monetary savings that can be achieved by using the different energy management schemes. For example, considering the femto priority scheme with 75 femtocells and the TX strategy, the yearly saving is about 1250 €. Considering a typical national network with about 50000 BSs for coverage, we can assume that about 20% of the BSs present cell overlap due to capacity issues in urban environment. Thus, the yearly saving in a national network can be roughly 2,4 millions of Euros. Moreover, if we consider the savings in terms of CO_2 emissions, this means a reduction of about 13000 tCO_2.

Finally, taking into account both CAPEX and OPEX and assuming that sleep modes are adopted at the BSs, we can conclude that (i) in the short term, the planning phase is the most crucial one from the cost point of view, due to the high CAPEX expenses; (ii) in the long term, significant savings can be achieved by adopting sleep modes at the BSs (due to OPEX reduction); (iii) among the different types of planning strategies that we have considered, the MP deployment (i.e., the one with many micro-BSs) is the most convenient one, and still some savings can be achieved with sleep modes at the micro-BSs.

8. Conclusions

This paper investigated algorithms and techniques that can be applied on cellular networks and provide traffic proportional power consumption. Three different planning strategies and BS management schemes were used to investigate potential savings in the network. Furthermore, the paper explored the effect of a heterogeneous network (femtocell layer) on BS management schemes by considering two different user-to-femto association rules. It was observed that the pseudo distributed management scheme together with femto priority association rule can provide important energy savings in the long run but also dynamic load control of the cellular network that is regarded as a fundamental element of future demand response services. Finally, it was proved that the femtolayer can improve QoS, coverage, and switch-off time intervals in the BS management scheme, providing more degrees of freedom to the mobile operator to adapt the power consumption in real time.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was funded from the European Union FP7/2007–2013 under Grant Agreement no. 257740 (Network of Excellence TREND). The authors would like to thank Dr. Anastastios Karousos for the outdoor channel predictions.

References

[1] W. Stark, H. Wang, A. Worthen, S. Lafortune, and D. Teneketzis, "Low-energy wireless communication network design," *IEEE Wireless Communications*, vol. 9, no. 4, pp. 60–72, 2002.

[2] W. Van Heddeghem, M. Deruyck, B. Puype et al., "Power consumption in telecommunication networks: overview and reduction strategies," *IEEE Communications Magazine*, vol. 49, pp. 62–69, 2011.

[3] 3GPP TR 32.826 v2.0.0 (2010-03), Study on Energy Savings Management (ESM), Release 9, 2010.

[4] 3GPP TR 36.902, E-UTRAN, Self-configuration and self-optimizing network use cases and solutions, (Rel 9), October 2009.

[5] A. P. Bianzino, C. Chaudet, D. Rossi, and J.-L. Rougier, "A survey of green networking research," *IEEE Communications Surveys and Tutorials*, vol. 14, no. 1, pp. 3–20, 2012.

[6] M. Gupta and S. Singh, "Greening of the Internet," in *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM '03)*, August 2003.

[7] L. Chiaraviglio, D. Ciullo, M. Meo, and M. A. Marsan, "Energy-efficient management of UMTS access networks," in *Proceedings of the 21st International Teletraffic Congress (ITC '09)*, Paris, France, September 2009.

[8] L. Chiaraviglio, D. Ciullo, G. Koutitas, M. Meo, and L. Tassiulas, "Energy-efficient planning and management of cellular networks," in *Proceedings of the 9th Annual Conference on Wireless On-Demand Network Systems and Services (WONS '12)*, pp. 159–166, January 2012.

[9] S. Kokkinogenis and G. Koutitas, "Dynamic and static base station management schemes for cellular networks," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM '12)*, Aneheim, Calif, USA, 2012.

[10] K. Samdanis, D. Kutscher, and M. Brunner, "Self-organized energy efficient cellular networks," in *Proceedings of the 21st IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC '10)*, pp. 1665–1670, September 2010.

[11] C. Peng, S. B. Lee, S. Lu, H. Luo, and H. Li, "Traffic-driven power saving in operational 3G cellular networks," in *Proceedings of the 17th Annual International Conference on Mobile Computing and Networking (MobiCom '11)*, Las Vegas, Nev, USA, September 2011.

[12] E. Oh, B. Krishnamachari, X. Liu, and Z. Niu, "Toward dynamic energy-efficient operation of cellular network infrastructure," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 56–61, 2011.

[13] Z. Niu, Y. Wu, J. Gong, and Z. Yang, "Cell zooming for cost-efficient green cellular networks," *IEEE Communications Magazine*, vol. 48, no. 11, pp. 74–79, 2010.

[14] M. Ismail and W. Zhuang, "Network cooperation for energy saving in green radio communications," *IEEE Wireless Communications*, pp. 76–81, 2011.

[15] M. Ajmone Marsan and M. Meo, "Energy efficient wireless Internet access with cooperative cellular networks," *Computer Networks*, vol. 55, no. 2, pp. 386–398, 2011.

[16] F. Richter, A. Fehske, P. Marsch, and G. Fettweis, "Traffic demand and energy efficiency in Heterogeneous cellular mobile radio networks," in *Proceedings of the International Conference on Wireless Communications (VTC '10)*, pp. 1–6, May 2010.

[17] K. Dufkova, M. Popovic, R. Khalili, J. V. Le Boudec, M. Bjelica, and L. Kencl, "Energy consumption comparison between macro-micro and public femto deployment in a plausible LTE network," in *Proceedings of the 2nd International Conference on Energy-Efficient Computing and Networking*, New York, NY, USA, 2011.

[18] G. Koutitas, A. Karousos, and L. Tassiulas, "Deployment strategies and energy efficiency of cellular networks," *IEEE Transactions on Wireless Communications*, vol. 11, no. 7, pp. 2552–2563, 2012.

[19] J. Lorincz, A. Capone, and D. Begusic, "Impact of service rates and base station switching granularity on energy consumption of cellular networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, article 342, 2012.

[20] S. Buzzi and D. Saturnino, "A game-theoretic approach to energy-efficient power control and receiver design in cognitive CDMA wireless networks," *IEEE Journal on Selected Topics in Signal Processing*, vol. 5, no. 1, pp. 137–150, 2011.

[21] O. Arnold, F. Richter, G. Fettweis, and O. Blume, "Power consumption modeling of different base station types in heterogeneous cellular networks," in *Proceedings of the Future Network and Mobile Summit*, Florence, Italy, June 2010.

[22] A. J. Fehske, F. Richter, and G. P. Fettweis, "Energy efficiency improvements through micro sites in cellular mobile radio networks," in *Proceedings of the IEEE Globecom Workshops (Gc Workshops '09)*, Honolulu, Hawaii, USA, December 2009.

[23] G. Auer, V. Giannini, I. Gódor et al., "Cellular energy efficiency evaluation framework," in *Proceedings of the 73rd IEEE Vehicular Technology Conference (VTC '11)*, May 2011.

[24] C. Khirallah, J. S. Thompson, and H. Rashvand, "Energy and cost impacts of relay and femtocell deployments in long-term-evolution advanced," *IET Communications*, vol. 5, no. 18, pp. 2617–2628, 2011.

[25] A. De Domenico, R. Gupta, and E. Calvanese Strinati, "Dynamic traffic management for green open access femtocell networks," in *Proceedings of the 75th IEEE Vehicular Technology Conference (VTC '12)*, pp. 1–6, May 2012.

[26] H.-F. Gerdees, *UMTS Radio Network Planning: Mastering Cell Coupling for Capacity Optimization*, Vieweg+Teubner Research, MRC, 2008.

[27] J. Zhang and G. de la Roche, *Femtocells: Technologies and Deployment*, John Wiley & Sons, 2010.

[28] J. M. Keenan and A. J. Motley, "Radio coverage in buildings," *British Telecom Technology Journal*, vol. 8, no. 1, pp. 19–24, 1990.

[29] 3GPP TSG-RAN WG4 #44-bis, "HNB and HNB-macro propagation models," 3GPP Report, 2007.

[30] J. Lorincz, T. Garma, and G. Petrovic, "Measurements and modelling of base station power consumption under real traffic loads," *Sensors*, vol. 12, no. 4, pp. 4281–4310, 2012.

[31] I. Haratcherev, C. Balageas, and M. Fiorito, "Low consumption home femto base stations," in *Proceedings of the 20th IEEE Personal, Indoor and Mobile Radio Communications Symposium (PIMRC '09)*, pp. 1–5, September 2009.

[32] G. Auer, V. Giannini, I. Godor et al., "Cellular energy efficiency evaluation framework," in *Proceedings of the 73rd IEEE Vehicular Technology Conference (VTC '11)*, 2011.

[33] M. Werner, M. Naden, P. Moberg et al., "Cost assessment of radio access network deployments with relay nodes," in *Proceedings of the IST Mobile and Wireless Communications Summit (IST '08)*, May 2008.

[34] International Energy Agency (IEA), "CO_2 emissions from fuel combustion: Highlights," Tech. Rep., 2009.

Hindawi Publishing Corporation
Journal of Electrical and Computer Engineering
Volume 2014, Article ID 806960, 11 pages
http://dx.doi.org/10.1155/2014/806960

Research Article

Facing the Reality: Validation of Energy Saving Mechanisms on a Testbed

Edion Tego,[1] Filip Idzikowski,[2] Luca Chiaraviglio,[3] Angelo Coiro,[3] and Francesco Matera[1]

[1] *Fondazione Ugo Bordoni, Viale del Policlinico 147, 00161 Rome, Italy*
[2] *TKN, Technische Universität Berlin, Einsteinufer 25, 10587 Berlin, Germany*
[3] *DIET, University of Rome, La Sapienza, Via Eudossiana 18, 00184 Rome, Italy*

Correspondence should be addressed to Luca Chiaraviglio; luca.chiaraviglio@diet.uniroma1.it

Received 7 November 2013; Revised 28 January 2014; Accepted 6 February 2014; Published 27 March 2014

Academic Editor: Vincenzo Eramo

Two energy saving approaches, called Fixed Upper Fixed Lower (FUFL) and Dynamic Upper Fixed Lower (DUFL), switching off idle optical Gigabit Ethernet (GbE) interfaces during low traffic periods, have been implemented on a testbed. We show on a simple network scenario that energy can be saved using off-the-shelf equipment not explicitly designed for dynamic on/off operation. No packet loss is experienced in our experiments. We indicate the need for faster access to routers in order to perform the reconfiguration. This is particularly important for the more sophisticated energy saving approaches such as DUFL, since FUFL can be implemented locally.

1. Introduction

Variation of traffic over day and night in backbone networks offers the opportunity to save energy by deactivating some network devices (or their parts) in low-demand hours. Various approaches have been proposed in the literature (see [1]) for choosing the devices to be deactivated. However there is little work done on the actual implementation validating the potential problems that energy saving schemes may introduce. The challenges (corresponding to steps schematically depicted in Figure 1) include (1) accurate monitoring of traffic data; (2) timely triggering network reconfiguration; (3) fast calculation of the desired (energy-efficient) network configuration; (4) the reconfiguration itself including signaling and the time needed to activate or deactivate the devices and potentially to reroute traffic. Consequently, issues such as network stability, increased delay, jitter, or even packet loss may occur in the network, which is particularly crucial in the backbone.

We validate the feasibility of implementation of two algorithms referred to as FUFL and DUFL [2–4] on a testbed. Our experiments show that it is possible to automatically and remotely switch on and off network interfaces in a dynamic manner using off-the-shelf equipment.

The structure of the paper is as follows. We provide an overview of the work related to experimental activities for green core networks in Section 2. The network scenario and methodology with algorithms' description are presented in Sections 3 and 4, respectively. Results are reported in Section 5. Eventually, Section 6 concludes this work.

2. Related Work

There is limited amount of related work dealing with implementation of energy saving mechanisms through selective on/off switching of network elements during periods of low load. To the best of our knowledge, it is basically limited to the following two activities.

2.1. MiDORi. Extensive work has been performed within the MiDORi (Multi (layer, path, and resources) Dynamically Optimized Routing) Network Technologies project [5]. The focus of the project is set on energy saving in the GbE network

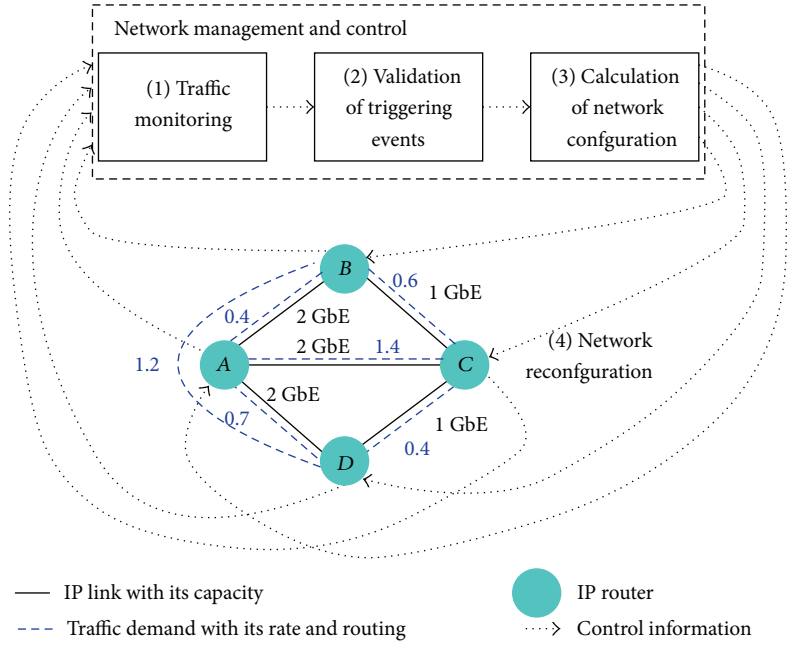


Figure 1: Steps needed for energy saving (traffic rates in Gbps, IP link capacities in number of GbE links).

by selectively powering off network interfaces under hop-limit and bandwidth constraints. Powering on/off the whole transit routers or their parts is also considered. Starting with [6], the authors propose a solution which effectively creates all on/off combinations of links in the network. The solution that can carry the whole traffic and consumes the lowest power (considering moving some virtual routers between network nodes in order to power off also nodes) is selected. Beeler's algorithm is compared with the proposed any-order pattern algorithm, which additionally guarantees the hop number of the path lower than a given maximum and disjoint multi-route link divergence for reliable communications. While the algorithms are centrally executed by the Path Computation Element (PCE), the authors consider also link on/off control protocols. Extensions to Generalized Multiprotocol Label Switching (GMPLS) (related to Open Shortest Path First (OSPF), Resources Reservation Protocol (RSVP), and Link Management Protocol (LMP)) are proposed.

Further publications related to MiDORi provide extensions of [6]. In particular, a prototype layer 2 (L2) switch is introduced in [7]. A depth-d algorithm is proposed and evaluated in a simulative way in [8]. The algorithm searches for the optimal configuration of the logical topology, where d determines the maximum number of links that are attempted to be switched off.

The following contributions are made in [9]. First, clear steps for the energy saving in the MiDORi architecture are described, that is, (1) traffic monitoring; (2) calculation of energy-efficient logical topology by PCE; (3) reconfiguration of the network. Second, experiments on a 6-node 7-link network using the depth-d algorithm are conducted, and results similar to the ones from [8] are presented together with the calculation times in the range of 0.01–10^3 s for networks with 10–100 nodes and d in the range of 1–4. In [9] the authors report also total current of the prototype L2 switches (2.765–2.831 A) and mention Ethernet Virtual Local Area Networks (VLANs) as the way of controlling traffic paths.

More details of the GbE L2 switch are provided in the block diagram presented in [10]. The switch can count traffic of each Label Switched Path (LSP) (VLAN) and each GbE link. The power consumption of the switch can be read via a command and a current meter. The presented switch has eight GbE links and is controlled remotely (power on/off state of each link and each fabric) using telnet via a Linux based control card which is one of the few parts constantly powered up. The authors demonstrated MiDORi on a fully meshed 6-node network testbed using the depth-1 algorithm with generic QoS restrictions. Six traffic generators/receivers were used; however the traffic assumptions were not detailed except for the fact that low traffic to high traffic ratio equals 1 : 5. The following steps are distinguished (extension from [9]). Step (1) is reading the traffic counters of each VLAN by the PCE and calculating average values. Step (2) is execution of the depth-1 algorithm at the PCE to obtain the logical topology and VLAN paths. Step (3-1) is powering on/off links in all switches (remotely by the PCE) according to the topology from Step (2). Step (3-2) is reconfiguration of the VLAN network topology according to the path calculation from Step (2).

Execution of Steps (1)–(3-2) is repeated every X minutes; however the authors do not report values assigned to X in [10]. Parallel and serial control of the switches were considered, with the parallel control taking significantly less time during both the traffic increase and traffic decrease (233.7–243.9 s versus 61.8–68.7 s, for serial versus parallel

control, resp.). The results show that the calculation of the logical topology (Step 2) takes marginal time (0.004–0.006 s). Duration of Steps (1), (3-1), and (3-2) takes 23.8–112.7 s in the serial control and 7.2–29.0 s in the parallel control.

The concept of Self-Organized Network (initially mentioned in [10]) is tackled in [11] using the depth-d algorithm again. The authors point out that the MiDORi GMPLS supports multiple layers, multiple paths, and multiple resources. They explain again the OSPF extension (relation between physical links and Traffic Engineering (TE) links), LMP extension (power on/off control function using the LMP ChannelStatus message with Ack and IP Control Channel always up), and RSVP extension (power control request in the Admin_Status object for LSP status flag). The authors mention the 16-port GbE switch, which they developed additionally to the 8-port switch presented in [10]. Demonstration on a 5-node 7-link network is performed showing that total switch power consumption can be reduced from 283.1 W to 276.1 W. The results of the reconfiguration times from [10] are also summarized in [11]. Additionally, the authors point out that their prototype switch does not support a "make-before-break" VLAN reconfiguration, and therefore data disruption occurs over the 29 s of VLAN reconfiguration. Eventually, the authors mention a MiDORi GMPLS optical switch, which they developed. It is also controlled via telnet by the PCE implemented on a small Linux box. The optical switch allows the authors to demonstrate the multilayer GMPLS signaling between a Lambda Switch Capable layer and a Layer 2 Switch Capable layer.

In [12], the authors show the energy consumption (in Wh without specifying the considered time period and details of the traffic data) on a 4-node full mesh network. The energy saving reaches up to 23.8%.

The experiments with the multilayer network using GbE switches and the optical switch from [11] are continued in [13] using the extension of the GMPLS. Namely, 4 Ethernet switches out of the considered 6 are connected to the optical switch. The demonstrated power saving (9.4 W corresponding to 6 ports) is low but shows that the MiDORi network technology is potentially feasible in a high speed and power consuming interface located in a large scale network environment.

Eventually, the results are summarized in [14], which directly extends [6]. It includes Beeler's algorithm, the any-order pattern algorithm, their simulative evaluation on the National Science Foundation (NSF) network loaded with uniformly generated internode traffic, the GMPLS extensions (OSPF, RSVP, and LMP), the 8-port GbE switch development, and the same results as in [10] (fully meshed 6-node network testbed).

The GMPLS extensions developed within the MiDORi project have been proposed to Internet Engineering Task Force (IETF) [15].

2.2. Experiments on the CARISMA Testbed. Researchers from Alcatel-Lucent Bell Labs and Universitat Politècnica de Catalunya (UPC) proposed the extension of GMPLS in [16] and performed experiments on the CARISMA testbed. More specifically, the authors of [16] propose to introduce a new bit "S" to the Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Path and Resv messages. The bit "S" used jointly with the already existing bit "A" differentiates the following states of an OptoElectronic (OE) device (such as colored line card, transponder, or regenerator): up, idle, down, and damaged (see also [17]).

The proposed extension has been evaluated on the CARISMA testbed available at UPC premises in Barcelona. The testbed was configured according to a Pan-European network composed of 16 nodes and 23 links with 10 bidirectional 100 Gbps wavelengths per link. 20 add/drop transponders and 10 regenerators were used at each node. The testbed was loaded with uniformly distributed lightpath requests according to the Poisson model with average holding time equal to 3 hours. The load was varied between 40 and 80 Erlang.

Differentiated provisioning of connection requests is considered in [18] for gold, silver, and best-effort traffic. In this case, the "A" and "S" bits of the ADMIN Status object of the RSVP-TE Path message are used in the following way: (i) "A" = 0 and "S" = 0 indicate the OE devices in the up state which must be used to allocate gold requests; (ii) "A" = 0 and "S" = 1 indicate the OE devices in idle state which are needed to allocate silver requests; (iii) "A" = 1 and "S" = 0 indicate the OE devices in down state which can be used to allocate best-effort requests. Differently to [16], availability of regenerators and wavelengths on links is disseminated over the network using the proposed extension of the GMPLS Open Shortest Path First-Traffic Engineering (OSPF-TE) protocol. A new sub-TLV (Type Length Value) (named TSP Status) is introduced in the OSPF-TE opaque Link State Advertisements (LSAs) containing the number of up, down, and idle transponders in a node (a regenerator corresponds to two transponders). This sub-TLV is inserted into a Node Information Top Level TLV (type 5, see Figure 1 of [18]). Using the OSPF-TE opaque LSAs, the PCE can populate its Traffic Engineering Database with wavelength and regenerator availability information, which is used for computation of end-to-end routes.

The same topology as in [16] is used for the experimental study on the CARISMA testbed [18]. The service class distribution is 20, 30, and 50% for gold, silver, and best-effort traffic, respectively. Different shares of resources are reserved for different classes of traffic. Prereservation of resources is implemented in the PCE to avoid contention of resources among different lightpaths under establishment. Results on the blocking ratio, number of OE devices in up/idle/down states, and power consumption per active LSP are reported.

The experimental activities on a testbed reported in [16, 18] were restricted to protocol information exchanges. Due to unavailability of transponders, the idle-up and down-up state transition times were assumed and not measured. The assumed transition times equal 20 ms and 60 s, respectively.

2.3. Our Contribution. We extend the related work described above in the following way. First, we implement the energy-saving schemes on the off-the-shelf equipment, demonstrating that the energy-saving is possible straightaway, even

without extending GMPLS, even though FUFL and DUFL can be used also in the GMPLS environment. Second, we explicitly consider triggering events for calculation of new network configuration and consequently potential network reconfiguration for energy saving. Third, we implement different energy-saving algorithms (FUFL and DUFL) than the ones implemented in the MiDORi and CARISMA testbeds. FUFL is particularly interesting for the network operators [4]. Fourth, we implemented the "make-before-break" mechanism. Eventually, we used different traffic schemes focusing on evaluation of deactivation of parallel GbE links constituting one logical link. While very little information is provided about traffic assumption in MiDORi studies, the traffic defined as the number of lightpath requests and not "bps to be transported" is used in [16, 18]. Finally, the work in [16, 18] does not consider routing of IP traffic over the logical topology.

3. Network Scenario

Even though the original methods FUFL and DUFL [2, 3] were proposed for IP-over-Wavelength Division Multiplexing (WDM) backbone networks, they can be applied also to other types of networks, as pointed out in [19, 20] for FUFL. Differently from [2, 3], optical GbE links are used instead of lightpaths in this work, which is determined by the testbed that we have access to.

3.1. Testbed. We used the testbed located at the Institute of Communications and Information Technology (Istituto Superiore delle Comunicazioni e delle Tecnologie dell'Informazione ISCOM) of the Italian Ministry of Economic Development. The testbed scheme is shown in Figure 2 [21]. The core part is composed of four Juniper M10/M10i routers (J1–J4) interconnected using 1 Gbps long haul optical links connecting Rome to Pomezia (total distance of 50 km). Three Cisco 3845 edge routers (C1–C3) are deployed at the access part of the network by means of GbE optical links. Finally, the testbed is completed with Gigabit Passive Optical Network (GPON) access networks composed of an Optical Line Terminal (OLT) and up to eight Optical Network Terminals (ONTs), offering a shared bandwidth equal to 1.244 Gbps. To guarantee an end-to-end minimum bandwidth in the backbone path, we use the technique described in [21] that allows us to assign a guaranteed bandwidth between two endpoints of the network by means of different tagging techniques, that is, VLAN and Virtual Private Local Area Network Service (VPLS).

Figure 3 shows the testbed in the configuration used in this work. A central Personal Computer (PC) is used for network management and control. It is connected directly over Fast Ethernet (FE) links to the IP router. Traffic monitoring is communicated by means of Simple Network Management Protocol (SNMP). Triggering of calculation of a new network configuration based on the monitored traffic as well as the calculation of the new network configuration itself is realized using bash scripting. Eventually, network reconfiguration is performed by logging via telnet into the IP routers and executing commands to perform rerouting and activation/deactivation of GbE interfaces.

Table 1: Traffic demands (bidirectional) in the network.

Traffic demand	Traffic type	Min (Gbps)	Max (Gbps)	Period (s)
A–H	Random	0.97	1	—
B–I	Sine-like	0	0.1	200
C–J	Sine-like	0	0.1	200
D–K	Random	0.97	1	—

3.2. Base Logical Topology. We focus on the core part consisting of one Juniper M10 and two M10i IP routers interconnected by 50 km of fiber cable into a bidirectional physical ring. The logical topology is composed of the IP routers interconnected by GbE optical links. All parallel GbE optical links between a node pair form a logical link. The logical topology is controlled in a centralized way [22].

We configure the testbed in order to obtain a simple scenario that allows demonstration of FUFL and DUFL. The base logical topology together with all traffic demands (source-target) is presented in Figure 4. The nodes *A–D* and *H–K* represent traffic generators and sinks attached to nodes *E* and *G*. The part representing a core network consists of the nodes *E*, *F,* and *G* interconnected by three logical links, each formed by two GbE optical links.

3.3. Traffic and Routing. Traffic and its routing over the base logical topology have been chosen in an artificial manner so that FUFL and DUFL operation can be demonstrated on the available testbed. The Ethernet Testing Platform Spirent SPT-3U, Anritsu MD1230B, and a Linux PC have been used to generate and terminate traffic. Two types of traffic are used: (i) random traffic with specified minimum and maximum values and (ii) sine-like traffic with specified minimum and maximum values, as well as period length in seconds. The sine-like traffic consists of halves of sine periods and of idle periods (as indicated in Figure 5(c)), which is determined by the traffic generators.

The maximum value of the sine-like traffic that our traffic generators can produce is 100 Mbps. Corresponding value for the random traffic is 1000 Mbps. A summary of the traffic inserted into the network is provided in Table 1. Traffic is generated in both directions.

The IP routing in the base network indicated in Figure 4 has been chosen so that all the logical links carry traffic and that the load exceeds the capacity of a single GbE optical link. It gives us the opportunity to see what happens with the traffic on a logical link when the whole logical link or just one out of two parallel GbE links is switched off in the low demand hour.

The inefficient utilization of the logical links is caused by the limitations of the traffic generators. This constitutes no obstacle for showing the operation of energy-saving approaches, but explicitly provides potential for switching off the GbE interfaces.

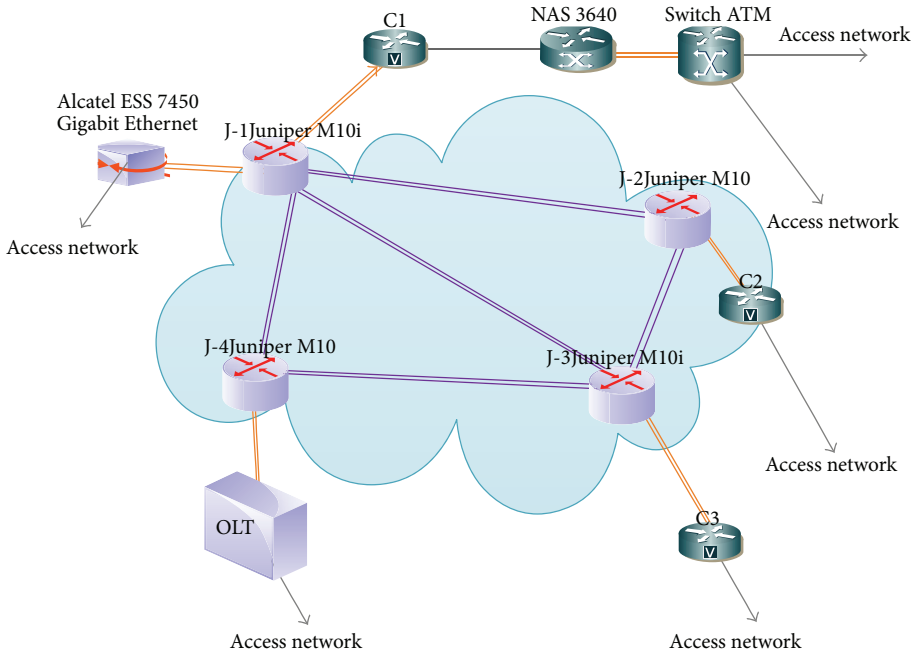


Figure 2: Testbed scheme located at Fondazione Ugo Bordoni (FUB) in Rome.

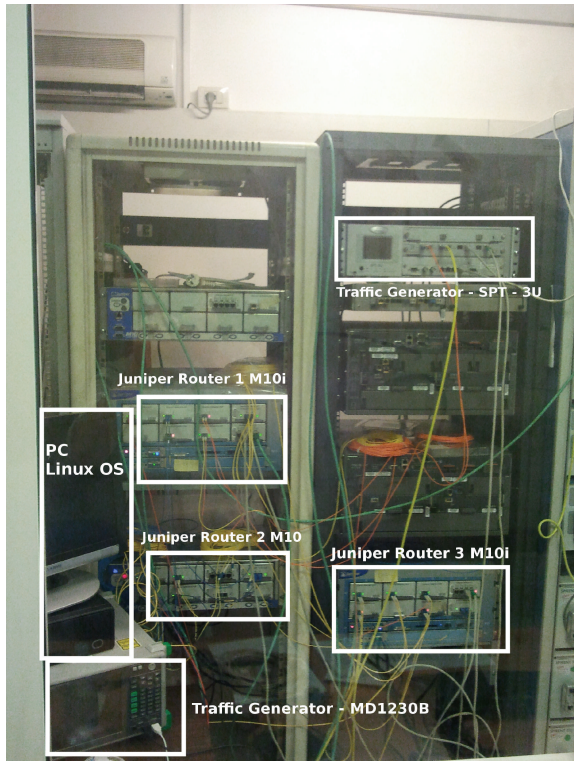


Figure 3: Photo of the testbed configured for the experiments.

3.4. Power Consumption. Power consumption of interfaces has been measured offline using Precision Power Analyzer N4L PPA2530 and a method similar to the one from [23]. Namely, power consumption of an IP router was measured twice, that is, when a GbE interface was active and when it was inactive (further measurements of power consumption of an IP router with all interfaces physically removed showed a difference of less than 0.5 W with respect to the router with an inactive GbE interface installed). The subtraction of these two values determined the power consumption of the

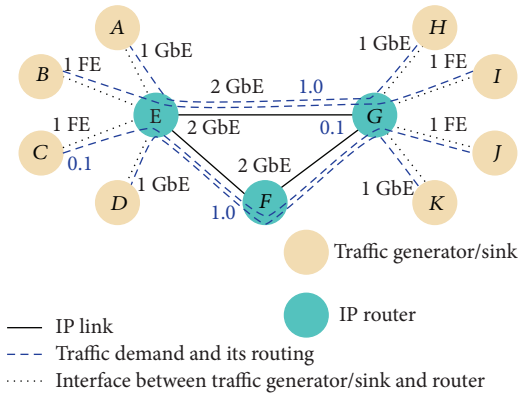


Figure 4: Base logical topology.

GbE interface (11.07 W for the M10 router and 8.9 W for the M10i router). Power consumption of Juniper's M10 and M10i routers with all interfaces shutdown equals (according to our measurements) 186.15 W and 112.5 W, respectively. The power consumption values reported above are lower than the ones reported in [23] for Cisco 7507 and in [11] for the MiDORi Ethernet switch.

4. Methods

We explain the methods used in each step needed for energy saving according to Figure 1.

4.1. Traffic Monitoring. The most intuitive approach to traffic monitoring is to perform the monitoring constantly. This approach is impossible in the simulative approaches performed so far, due to unavailability of input traffic data sets originating from measurements and covering traffic between all node pairs in the network over sufficiently long period of time with sufficient time granularity (IP packet arrival/departure level). Experimental activity overcomes this limitation. However, there is still some level of freedom in setting the traffic monitoring in a digital system such as a telecommunication network (testbed), namely, the time period T_M over which the constantly monitored traffic data is averaged and provided for evaluation triggering calculation of a new network configuration.

4.2. Validation of Triggering Events. The second level of freedom determines the frequency at which the events triggering calculation of network configuration are validated. The related time period is denoted as T_L and should be set as small as possible in the digital system in order to mimic constant validation of triggering events. Please note that T_L is different than T_M. The latter determines the history which is taken into account when validating triggering events, while the former determines the frequency at which the validation is performed.

With reference to Figure 1, the "traffic monitoring" block provides with periodicity T_L input data (measure of the traffic load experienced on each logical link) for the "validation of triggering events" block. Thus, starting from the time instant t_1, corresponding to the first validation, the nth validation is performed at time $t_n = t_1 + (n-1)T_L$. Each validation is performed according to the measure of the average traffic load experienced during a period T_M. Specifically, the measure provided at time t_n corresponds to the average traffic load during the period $(t_n - T_M, t_n)$. Notice that choosing a small value for T_L allows a prompt reaction to changing traffic conditions, which is particularly important during increasing traffic trend in order not to experience congestion within the network. On the other hand, the system should be as much stable as possible and not follow all tiny variations of traffic. For this reason, traffic load measures are provided as averages over the period T_M, which should be chosen sufficiently long so as to hide very high frequency traffic variations.

Calculation of new network topology is triggered by violation of thresholds (W_A and W_D for FUFL and W_L and W_H for DUFL) as explained in detail in the following subsection. This step also takes into account the stability issue by including a hysteresis cycle within the threshold mechanism.

4.3. Calculation of Network Configuration. We focus on two classes of approaches to calculation of network configuration [2, 3], namely, FUFL and DUFL.

4.3.1. FUFL. The first class is very simple and attractive for network operators [4]. It is fully distributed and involves neither changing of IP routing nor changing of the connectivity of the logical topology. The load on each GbE link constituting the logical link is monitored. A GbE link is switched off when load on the previous parallel GbE link goes down below W_D. It is switched on again when the load on the previous parallel GbE link goes above W_A. W_D and W_A are defined as utilization of a GbE link. The explanation above assumes bin-packing of traffic in parallel links [19, 20]. We have not verified the load-balancing mechanisms used in the Juniper routers. If other packing (load-distribution) strategies are used, traffic on the GbE link to be switched off is shifted to other active parallel GbE links. According to [24, 25], link aggregation implementation in Juniper routers uses the same load-balancing algorithm as that used for per-packet load balancing; that is, the router sends successive data packets over paths without regard to individual hosts or user sessions. It uses the round-robin method to determine which path each packet takes towards its target.

4.3.2. DUFL. The second class DUFL is more complex, as it allows changing of IP routing, which in turn may increase the number of idle interfaces in the network and lead even to switching off whole logical links. A logical link nonexisting in the base network cannot be established though. There are many algorithms which fall into the class of DUFL (see, e.g., DAISIES [26] and Least Flow Algorithm [27]); however their comparison is out of the scope of this paper. The power savings in the simple base network topology from Figure 4

would be basically identical. A simulative comparison on a larger network is available in [1].

For the sake of this work, we assume the following implementation of DUFL. The decision about an attempt to reroute traffic with the aim of deactivation or activation of a logical link is triggered by violation of the thresholds W_L and W_H, respectively. Both W_L and W_H are defined as utilization of a logical link. The traffic demands routed via *E-F-G* are attempted to be rerouted to link *E–G* if aggregated demand on the logical links *E-F* and *F-G* goes below W_L. Analogical rerouting attempt is performed when load of the logical link *E–G* goes below W_L.

Idle logical links with optical interfaces are switched off. The original logical topology and routing (Figure 4) are restored when W_H is violated on any logical link.

4.4. Network Reconfiguration. The last step concerns the application of the newly computed network configuration in network devices. To perform this step, the management system opens a telnet session on routers which need their configurations to be changed and applies the needed changes. Specifically, routing is changed and network interfaces are switched on/off according to the computed network configuration. We ensure that rerouting is performed before a logical link is released when load decreases and after a logical link is established when load increases.

5. Results

We parameterize the methods described in Section 4 in the following way. The thresholds are assigned with the following values: $W_D = 0.977$ and $W_A = 0.985$ for FUFL and $W_L = 0.4885$ and $W_H = 0.9925$ for DUFL. Both T_M and T_L are set to 10 s. The chosen values are determined by the generated traffic characteristic and for the sake of demonstration of the power saving approaches. Traffic variations close to the threshold values allow us to verify the methods without waiting long (corresponding to diurnal variation of traffic).

As for the setting of T_L (the time period between two successive validations of triggering events), we had to mind the time needed for reconfiguration (Step 4 in Figure 1). Specifically, T_L should be longer than the time needed for network reconfiguration, denoted as Δ_{Step4}. The reason is that two concurrent attempts to switch on/off a network interface could be undertaken otherwise. Due to the nonnegligible values of Δ_{Step4} experienced in our experiments, we decided to overcome this problem by not performing the validation of triggering events during network reconfiguration. In this way, it was possible to keep T_L lower than Δ_{Step4} without experiencing any concurrent attempts to switch on/off an interface.

Figures 5(a) and 5(b) report the total power consumption and the power saving for the testbed running FUFL and DUFL on the logical topology, respectively. For clarity, the first 600 seconds are reported. The total power consumption corresponds to power consumed by all active GbE interfaces together with the routers according to the data from Section 3.4. Power consumption varies more frequently with DUFL than with FUFL, since our implementation of DUFL is more aggressive in turning off the network interfaces—it attempts to switch off the whole logical links. This in turn produces in general higher power saving compared to FUFL. The difference is minor due to the simple 3-node base logical topology and IP routing schemes that we use for this demonstration (see Section 3).

Figures 5(c) and 5(d) report the monitored traffic on logical links when FUFL and DUFL are applied, respectively. The figures report also the sine-like traffic injected to the network. As expected, all the logical links are always utilized with FUFL, and therefore each link has always aggregated traffic around 1 Gbps. On the contrary, the utilization of the links frequently changes with DUFL, since this algorithm reroutes the traffic and powers off the entire logical links. Therefore traffic on each logical link has a strong fluctuation between 0 and 2 Gbps.

The time Δ_{Step3} needed for calculation of network configuration (Step 3 in Figure 1) takes 0.15 s for both FUFL and DUFL. We measured also the time needed for network reconfiguration (Step 4 in Figure 1). It consists of (i) time consumed by telnet (opening a session), Δ_{telnet}: 11.54 s; (ii) time needed to power on a GbE interface, Δ_{activate}: 0.01 s; (iii) time needed to power off a GbE interface, $\Delta_{\text{deactivate}}$: 0.01 s. We neglect the time that is needed to perform the rerouting in DUFL.

We point out that the time values that we obtained are comparable with the ones reported in [10, 11], even though the testbeds differ significantly.

The overall duration of Step 4 for a network consisting of a set of nodes V is calculated according to

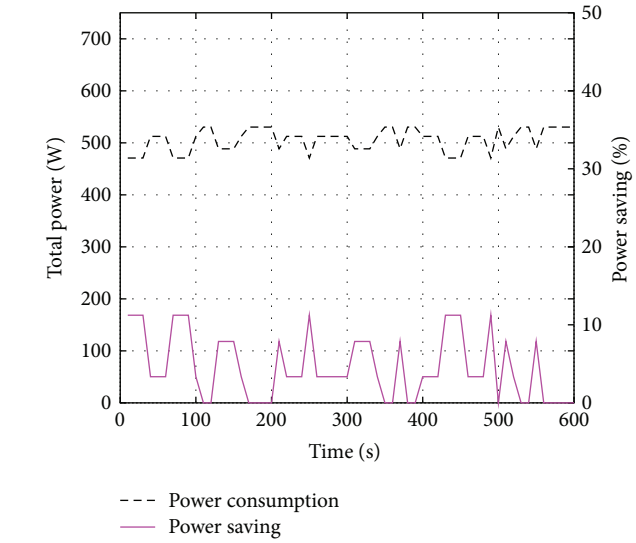
$$\Delta_{\text{Step4}} = \sum_{i \in V} \left(x_i \cdot \Delta_{\text{activate}} + y_i \cdot \Delta_{\text{deactivate}} \right) + z \cdot \Delta_{\text{telnet}}, \quad (1)$$

where x_i and y_i denote the number of activated and deactivated interfaces at node $i \in V$, respectively, and z denotes the number of nodes (routers) that need to be accessed one after another.

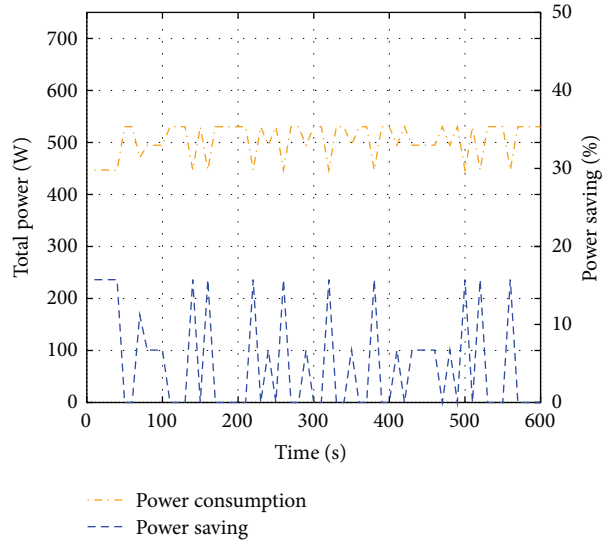
Clearly, in an operational network, the time Δ_{Step4} required for the reconfiguration should be limited. In our case, it takes quite a lot of time to open a telnet session in order to reconfigure a router. If telnet authentication is done manually, the time to open a telnet session decreases a lot for the routers under consideration, depending on the complexity of the password. For example, the time that we measured with a manual authentication and a simple password of 6 characters was only 2.5 seconds. This amount of time depends on the implementation of telnet and on the operating system of the device. In our case the telnet sessions are opened one after another in order to perform the configuration of the interfaces for each node. In such a case, $z = \sum_{i \in V} z_i$, where z_i determines the need to access node $i \in V$; namely,

$$z_i = \begin{cases} 0, & \text{if } x_i + y_i = 0, \\ 1, & \text{otherwise.} \end{cases} \quad (2)$$

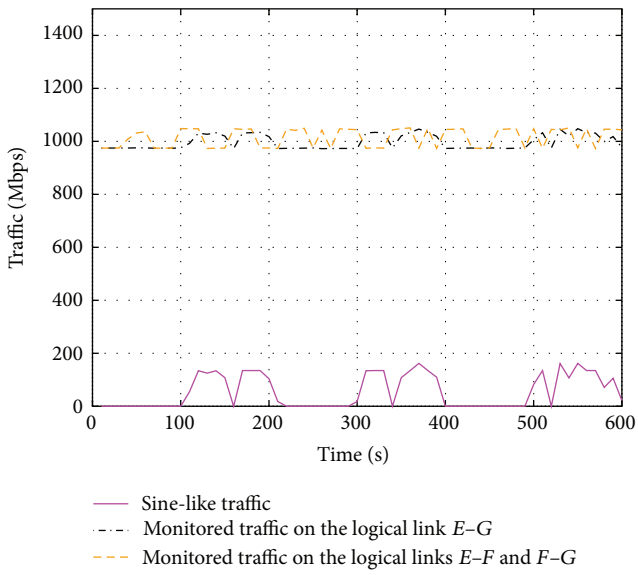
Note that z can be at most reduced to 1 if routers can be accessed in parallel. We show the time taken by network



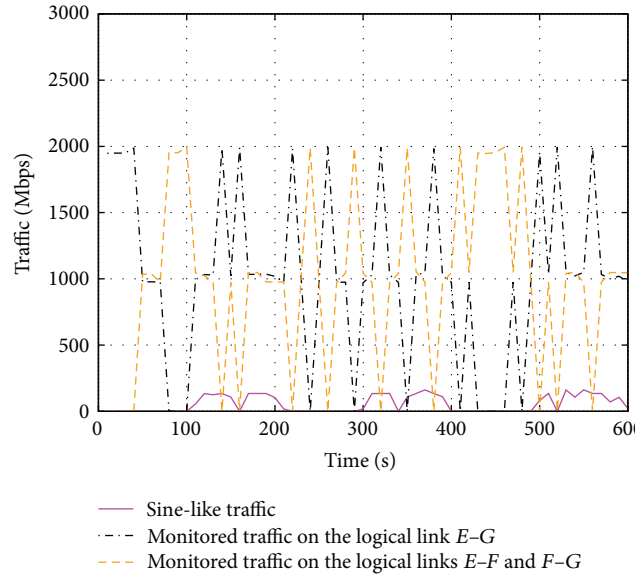
(a) FUFL: total power consumption (left y-axis) and power saving (right y-axis)



(b) DUFL: total power consumption (left y-axis) and power saving (right y-axis)



(c) FUFL: monitored traffic



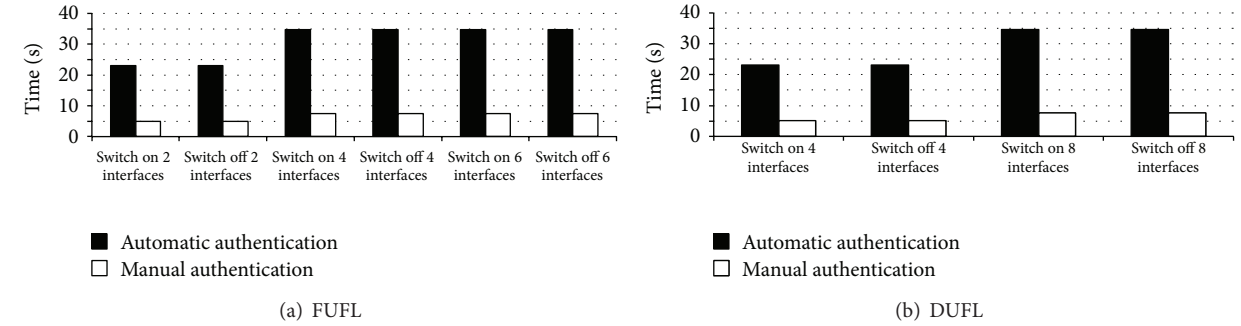
(d) DUFL: monitored traffic

Figure 5: FUFL and DUFL results on the testbed (mind different y-scales in (c) and (d)).

reconfiguration Δ_{Step4} in Figure 6 for FUFL and DUFL and for both the parallel and consecutive access of routers (automatic and manual authentication, resp.). Clearly, the required time to switch on/off the devices increases, when the number of affected interfaces is increased, since more routers need to be accessed and configured in our scenario. The theoretical time taken to switch on/off interfaces when manual authentication is assumed is significantly lower than the time measured during our experiments with automatic authentication. We think that this issue of long automatic authentication should be easily overcome in the operational network. Furthermore, local implementation of FUFL on routers would avoid this problem too.

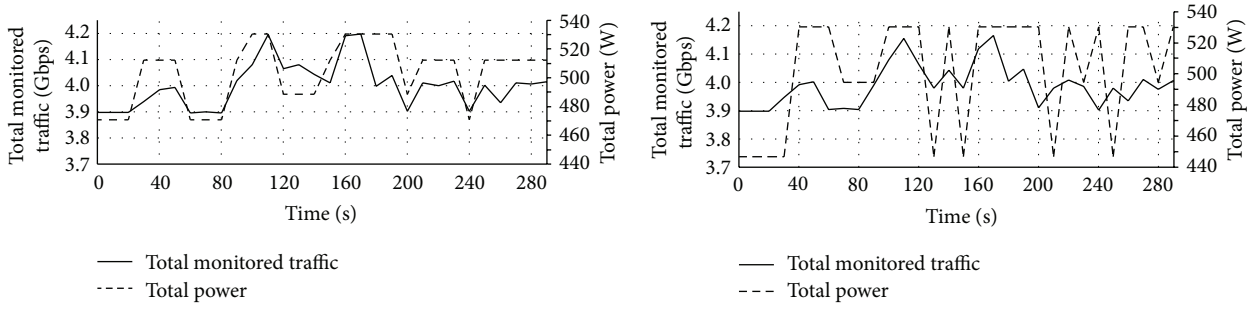
In Figure 7 we can observe the trends of the total monitored traffic and total power consumed by the network over time. For clarity we limit the timescale to the first 290 s. We can clearly see the direct impact of the total monitored traffic on the total power. In particular, the power tends to increase when the traffic increases, meaning that the algorithms are able to correctly react to the traffic variation. We can observe again that with DUFL the power varies more frequently, suggesting that the number of interfaces that are switched on/off frequently varies.

To give more insight, we have collected the events that occur in the network. In particular, every time that one of the thresholds is violated (Step 2 in Figure 1), one of the actions



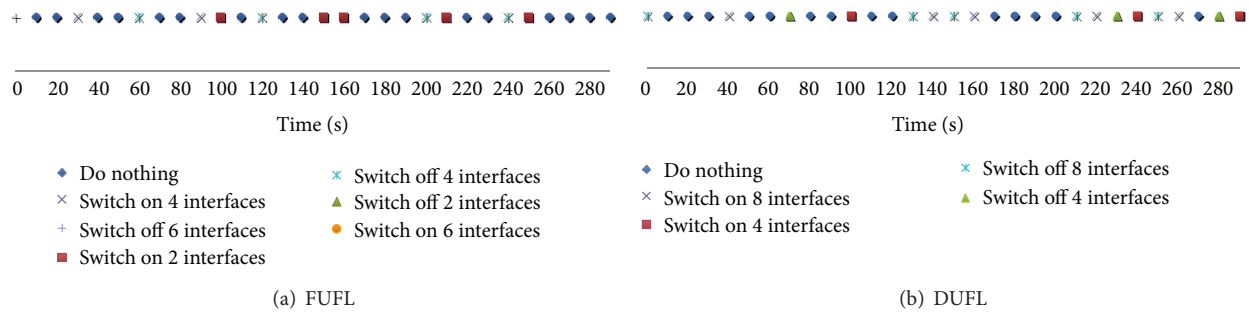
(a) FUFL (b) DUFL

Figure 6: Time taken by network reconfiguration (Δ_{Step4}).



(a) FUFL (b) DUFL

Figure 7: Total monitored traffic and total power over time (mind different y-scales).



(a) FUFL (b) DUFL

Figure 8: Occurring action over time.

is taken according to the calculated (with FUFL or DUFL) network configuration (Step 3 in Figure 1). The occurrence of such events is reported in Figure 8. We show the first 290 s on the time axis in correspondence with Figure 7. Interestingly, we can observe that with FUFL (Figure 8(a)) most of the events do not require a change in the current network configuration, while with DUFL (Figure 8(b)) the configuration frequently varies over time. Additionally, the number of switched off interfaces is higher with DUFL than with FUFL (as expected).

Finally, we observed minor increase of end-to-end packet delay (up to 30 ms during reconfiguration) and no packet loss based on the monitored traffic. Thus, we can conclude that our solutions are able to save energy while not deteriorating the quality of service for users in this simple scenario.

6. Conclusion

We demonstrated the operation of energy saving approaches FUFL and DUFL on a testbed with optical GbE interfaces. We showed that it was possible to dynamically adapt the network configuration to the changing load without losing traffic and with a minor increase in the packet delay. Moreover, we demonstrated that it was feasible to automatically and remotely activate and deactivate interfaces of commercial devices available today. We experienced relatively long time to reconfigure the network, which depends on how routers are accessed. We believe that this issue can be easily overcome in the operational networks using future devices designed for green networking. Furthermore, FUFL results can be directly translated to bigger networks, because this mechanism is local. Experimental validation of DUFL approaches is more

challenging, since no testbed of a large size is available to us. Results of simulative studies on larger networks than our scenario can be found in [1, 4, 26, 28].

In general, the mesh degree of the network influences power savings more than the network size [29, 30]. This is due to the fact that the mesh degree determines the number of possibilities to reroute traffic. Regarding the different methods for calculation of network configuration (Step 3 in Figure 1), our big picture study [1] showed that there is a noticeable difference between the simple local method FUFL and the more complex methods. However, power saving should be evaluated together with other evaluation criteria such as impact on QoS, network knowledge, or protection consideration (see Table 2 in [1]).

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under Grant agreement no. 257740 (Network of Excellence "TREND").

References

[1] F. Idzikowski, E. Bonetto, L. Chiaraviglio et al., "TREND in energy-aware adaptive routing solutions," *IEEE Communications Magazine*, vol. 51, no. 11, pp. 94–104, 2013.

[2] F. Idzikowski, S. Orlowski, C. Raack, H. Woesner, and A. Wolisz, "Saving energy in IP-over-WDM networks by switching off line cards in low-demand scenarios," in *Proceedings of the 14th International Conference on Optical Networking Design and Modeling (ONDM '10)*, Kyoto, Japan, February 2010.

[3] F. Idzikowski, S. Orlowski, C. Raack, H. Woesner, and A. Wolisz, "Dynamic routing at different layers in IP-over-WDM networks Maximizing energy savings," *Optical Switching and Networking*, vol. 8, no. 3, pp. 181–200, 2011.

[4] F. Idzikowski, L. Chiaraviglio, R. Duque, F. Jimenez, and E. Le Rouzic, "Green horizon: looking at backbone networks in 2020 from the perspective of network operators," in *Proceedings of the International Conference on Communications (ICC '13)*, Budapest, Hungary, June 2013.

[5] MiDORi Network Technologies Project (MiDORi), 2009, http://midori.yamanaka.ics.keio.ac.jp.

[6] N. Yamanaka, S. Shimizu, and G. Shan, "Energy efficient network design tool for green IP/Ethernet networks," in *Proceedings of the 14th International Conference on Optical Networking Design and Modeling (ONDM '10)*, Kyoto, Japan, February 2010.

[7] N. Yamanaka, H. Takeshita, S. Okamoto, and S. Gao, "[Invited] MiDORi: Energy efficient network based on optimizing network design tool, remote protocol and new layer-2 switch," in *Proceedings of the 9th International Conference on Optical Internet (COIN '10)*, Jeju, South Korea, July 2010.

[8] H. Yonezu, S. Gao, S. Shimizu et al., "Network power saving topology calculation method by powering off links considering QoS," in *Proceedings of the 15th OptoElectronics and Communications Conference (OECC '10)*, pp. 586–587, Sapporo, Japan, July 2010.

[9] H. Yonezu, K. Kikuta, D. Ishii, S. Okamoto, E. Oki, and N. Yamanaka, "QoS aware energy optimal network topology design and dynamic link power management," in *Proceedings of the 36th European Conference and Exhibition on Optical Communication (ECOC '10)*, Torino, Italy, September 2010.

[10] H. Takeshita, Y. Oikawa, H. Yonezu, D. Ishii, S. Okamoto, and N. Yamanaka, "Demonstration of the self organized dynamic link power management by "MiDORi" energy optimal network topology design engine," in *Proceedings of the Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC '11)*, Los Angeles, Calif, USA, March 2011.

[11] S. Okamoto, Y. Nomura, H. Yonezu, H. Takeshita, and N. Yamanaka, "GMPLS-enabled, energy-efficient, self-organized network: MiDoRi," in *Proceedings of the Asia Communications and Photonics Conference and Exhibition (ACP'11)*, Shanghai, China, November 2011.

[12] Y. Nomura, H. Yonezu, D. Ishii, S. Okamoto, and N. Yamanaka, "Dynamic topology recon guration for energy efficient network with link power control: MiDORi," in *Proceedings of the World Telecommunications Congress (WTC'12)*, Yokohama, Japan, March 2012.

[13] Y. Nomura, H. Yonezu, and D. Ishii, "Dynamic topology reconfiguration for energy efficient multi-layer network using extended GMPLS with link power control," in *Proceedings of the Optical Fiber Communication Conference and Exposition (OFC/NFOEC '12)*, Los Angeles, Calif, USA, March 2012.

[14] H. Takeshita, N. Yamanaka, S. Okamoto, S. Shimizu, and S. Gao, "Energy efficient network design tool for green IP/Ethernet networks," *Optical Switching and Networking*, vol. 9, no. 3, pp. 264–270, 2012.

[15] S. Okamoto, "Requirements of GMPLS extensions for energy efficient traffic engineering," 2013, http://tools.ietf.org/html/draft-okamoto-ccamp-midori-gmpls-extension-reqs-02.

[16] A. Morea, S. Spadaro, O. Rival, J. Perelló, F. Agraz, and D. Verchere, "Power management of optoelectronic interfaces for dynamic optical networks," in *Proceedings of the 37th European Conference on Optical Communication and Exhibition (ECOC '11)*, Geneva, Switzerland, September 2011.

[17] A. Morea, J. Perelló, S. Spadaro, D. Verchère, and M. Vigoureux, "Protocol enhancements for "greening" optical networks," *Bell Labs Technical Journal*, vol. 18, no. 3, pp. 211–230, 2013.

[18] A. Morea, J. Perelló, F. Agraz, and S. Spadaro, "Demonstration of GMPLS-controlled device power management for next generation green optical networks," in *Proceedings of the Optical Fiber Communication*, Los Angeles, Calif, USA, March 2012.

[19] L. Liu and B. Ramamurthy, "Rightsizing bundle link capacities for energy savings in the core network," in *Proceedings of the 54th Annual IEEE Global Telecommunications, Exhibition and Industry Forum (GLOBECOM '11)*, Houston, Tex, USA, December 2011.

[20] L. Lin and B. Ramamurthy, "A dynamic local method for bandwidth adaptation in bundle links to conserve energy in core networks," *Optical Switching and Networking*, vol. 10, no. 4, pp. 481–490, 2013.

[21] A. Valenti, A. Rufini, S. Pompei et al., "QoE and QoS comparison in an anycast digital television platform operating on passive optical network," in *Proceedings of the Telecommunications*

Network Strategy and Planning Symposium (NETWORKS '12), Rome, Italy, October 2012.

[22] L. Rea, S. Pompei, A. Valenti, F. Matera, C. Zema, and M. Settembre, "Quality of Service control based on Virtual Private Network services in a Wide Area Gigabit Ethernet optical test bed," *Fiber and Integrated Optics*, vol. 27, no. 4, pp. 301–307, 2008.

[23] J. Chabarek, J. Sommers, P. Barford, C. Estan, D. Tsiang, and S. Wright, "Power awareness in network design and routing," in *Proceedings of the 27th IEEE Communications Society Conference on Computer Communications (INFOCOM '08)*, pp. 1130–1138, Phoenix, Arizona, USA, April 2008.

[24] Juniper, "JUNOS Software Interfaces and Routing Configuration Guide," 2010, http://www.juniper.net/techpubs/software/junos-security/junos-security10.2/junos-security-swconfig-interfaces-and-routing/junos-security-swconfig-interfaces-and-routing.pdf.

[25] Cisco, "Per-Packet Load Balancing, IOS Release 12.2(28)SB Guide," 2006, http://www.cisco.com/en/US/docs/ios/12_0s/feature/guide/pplb.pdf.

[26] A. Coiro, M. Listanti, A. Valenti, and F. Matera, "Energy-aware traffic engineering: a routing-based distributed solution for connection-oriented IP networks," *Computer Networks*, vol. 57, no. 9, pp. 2004–2020, 2013.

[27] L. Chiaraviglio, M. Mellia, and F. Neri, "Minimizing ISP network energy cost: formulation and solutions," *IEEE/ACM Transactions on Networking*, vol. 20, no. 2, pp. 463–476, 2012.

[28] E. Bonetto, L. Chiaraviglio, F. Idzikowski, and E. Le Rouzic, "Algorithms for the multi-period power-aware logical topology design with reconfiguration costs," *Journal of Optical Communications and Networking*, vol. 5, no. 5, pp. 394–410, 2013.

[29] W. Van Heddeghem, F. Musumeci, F. Idzikowski et al., "Power consumption evaluation of circuit-switched versus packet-switched optical backbone networks," in *Proceedings of the OnlineGreenComm*, October 2013.

[30] L. Chiaraviglio, D. Ciullo, M. Mellia, and M. Meo, "Modeling sleep mode gains in energy-aware networks," *Computer Networks*, vol. 57, no. 15, pp. 3051–3066, 2013.

Hindawi Publishing Corporation
Journal of Electrical and Computer Engineering
Volume 2014, Article ID 683408, 12 pages
http://dx.doi.org/10.1155/2014/683408

Research Article

Design of a Traffic-Aware Governor for Green Routers

Alfio Lombardo, Vincenzo Riccobene, and Giovanni Schembra

Dipartimento di Ingegneria Elettrica, Elettronica e Informatica (DIEEI), University of Catania, Viale A. Doria 6, 95125 Catania, Italy

Correspondence should be addressed to Giovanni Schembra; schembra@dieei.unict.it

Received 7 November 2013; Revised 11 January 2014; Accepted 15 January 2014; Published 12 March 2014

Academic Editor: Vincenzo Eramo

Today the reduction of energy consumption in telecommunications networks is one of the main goals to be pursued by manufacturers and researchers. In this context, the paper focuses on routers that achieve energy saving by applying the frequency scaling approach. The target is to propose an analytical model to support designers in choosing the main configuration parameters of the Router Governor in order to meet Quality of Service (QoS) requirements while maximizing energy saving gain. More specifically, the model is used to evaluate the input traffic impacts on the choice of the active router clock frequencies and on the overall green router performance. A case study based on the open NetFPGA reference router is considered to show how the proposed model can be easily applied to a real case scenario.

1. Introduction

In the last decade, new requirements are appearing in telecommunications network design and management deriving from the fact that the global Internet, with its energy consumption of about 8% of the global production, is becoming one of the most important energy consumers in the world [1]. Today's most telecommunications networks are provisioned for worst-case or busy-hour load, and this load typically exceeds their long-term utilization by a wide margin; moreover, as shown in [2], current network nodes have a power consumption that is practically constant and does not depend on the actual traffic load they face. The implication of these factors is that most of the energy consumed in networks today is wasted [3]. A nonmarginal side effect of high-energy dissipation is the increment of the temperature of the places where network devices reside, with a consequent further waste of energy used by cooling machines to maintain the temperature of the local environment constant.

For this reason, addressing energy efficiency in the Internet is receiving considerable attention in the literature today [4–10] and many research projects are working on this topic (see, e.g., [11–13]). The novel approach for networking means that, besides typical performance parameters as, for example, throughput, latency, and packet loss probability, amount of consumed energy starts to be one of the most important factors of network design and operation.

For the above reasons, some novel hardware devices, the so-called "green routers", are expected in the near future to allow different power states [14] according to the input traffic. A lot of work was done in the past, focusing on the definition of power management techniques, like, for example, the static techniques described in [6–8], and the adaptive policy proposed in [9]. Two approaches have been proposed to reduce energy consumption in network components [5]. The first is based on putting network components in sleeping state during idle intervals, reducing energy consumed in the absence of traffic. The second one is based on adapting the rate of network operations to the offered workload. Rate adaptation in particular is usually achieved by scaling the processing power according to the data rate the router has to manage; at this purpose, the clock frequency driving the router processes can be modified according to the input data rate [10]. The energy aware techniques to be used in a green router depends on a number of factors, including the role of the router in the network, the profile of incoming traffic, and the hardware complexity. Other aspects that have to be

considered are the related costs with respect to the energy we can potentially save, and the Quality of Service (QoS) we want to guarantee to the users [15]. The user notices also that different techniques and architectures have been proposed in the literature in order to provide frequency scaling capabilities to networking devices, like, for example, [16, 17].

With all this in mind, the paper focuses on routers that achieve energy saving by applying the frequency scaling approach [17]. The target is to extend the proposed model of a green router introduced by the same authors in [18, 19] to support designers in choosing system parameters in order to meet QoS requirements while maximizing energy saving gain. The paper starts from the observation that each modification of the operating clock frequency causes some QoS degradation in terms of packet loss, delay, or energy waste, according to the particular implementation of the router. For this reason, the best tradeoff between energy saving and QoS performance could be achieved by using a set of clock frequencies that is a timely chosen subset of all the clock frequencies supported by the router CPU. However, the choice of the particular subset, that is, both the number of frequencies and which frequencies among all the available ones, is strongly related to the input traffic, and specifically its mean value, its variance, and its autocorrelation. For example, it is not befitting to use clock frequencies that manage bit rate values close to the mean value of the input traffic bit rate. Moreover, it is better to avoid frequencies that are very close to each other if the traffic is low correlated. With the aim of choosing the set of frequencies and deciding the best clock frequency at runtime, a Router Governor is introduced. An additional parameter, in the following referred to as δ, is introduced to control the frequency change rate, with the aim of matching the given QoS requirements. Starting from the Router Governor architecture defined in [18, 19], defined to support only two clock frequencies, a general Router Governor is proposed to work in routers with any number of clock frequencies. A new multidimensional discrete-time Markov model is presented to capture the behavior of the proposed Governor. Since, as mentioned so far, each frequency switch is characterized by a given cost, the model is used to evaluate how the input traffic impacts the choice of the active clock frequencies and on the overall green router performance. A case study based on the open NetFPGA Open Router [20] is considered to show how the proposed model can be easily applied to a real case scenario. More specifically, the paper uses the green NetFPGA Reference Router proposed by the same authors in [18, 19] that leverages on the facility of the NetFPGA platform to reduce the clock rate by changing the value of an ad-hoc hardware register. Loss probability and energy saving gain are considered as QoS metrics.

The paper is structured as follows. Section 2 introduces the reference router architecture and the proposed policy. Section 3 describes the Markov model of the considered system. Section 4 derives of the main performance parameters. All the results of our analysis are shown in Section 5, which describes the proposed case study. Finally, Section 6 ends the paper with authors' conclusions and future directions.

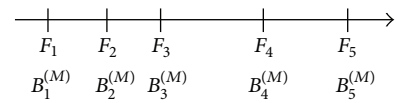


Figure 1: Set of clock frequencies implemented by the router, and relative maximum supported bitrates.

2. A Traffic-Aware Governor for Green Routers

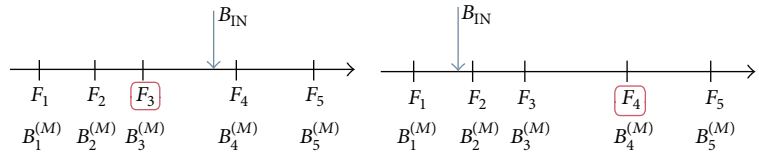
In this section we describe the system which is the focus of this paper. It is a Governor for green routers that implement frequency scaling [19] to save energy when the input traffic load is low. Frequency scaling, a capability available in many routers today, is the possibility of changing the core clock frequency in a set of values to dynamically scale the energy consumption of the device. The base problem of this approach is that if on the one hand the device power consumption is reduced using lower clock frequencies with respect to the highest one, on the other hand such a decision can deteriorate the router performance. For example, in the green implementation of the NetFPGA Reference Router [18], clock frequency switches cause a temporary block of the router, and therefore all the incoming packets during these intervals are lost. Other routers, although with different hardware architecture and implementation, behave at the same way: at each clock frequency variation they present a QoS degradation, in terms of either loss probability, delay, and/or energy consumption peaks.

Starting from the above considerations, the approach proposed in this paper, which aims at finding the best tradeoff between energy efficiency and QoS, is very general since it can be used to limit such a router QoS degradation by only changing the particular target QoS parameter (e.g., loss probability, mean delay, or energy consumption during the switching periods).

In order to manage frequency switches maintaining QoS acceptable while decreasing energy consumption, we introduce a Router Governor, that is, an entity which implements a router management policy to change the clock frequency of the router CPU. In the following, QoS is defined by the following parameters: packet loss, mean delay, and energy waste during frequency switching intervals.

Let us note that other traditional QoS parameters characterizing the router, like, for example, packet loss probability for output queue overflow and queuing delay, are not considered here because they are not altered by the presence of our Router Governor.

Let $\overline{\Phi}$ be the set of clock frequencies supported by the router CPU, and let F_i be the generic ith CPU clock frequency. For the sake of simplicity, we sort frequencies in such a way that $F_i < F_{i+1}$. Let us indicate the maximum bit rate that can be supported with no loss when the CPU is working at the frequency F_i as $B_i^{(M)}$. These values are sketched in Figure 1.



(a) Example for Rule 1 (b) Example for Rule 2

Figure 2: Set of clock frequency implemented by the router, and relative maximum supported bitrates.

An important observation that is at the basis of our approach is that the greater the cardinality of $\overline{\Phi}$, that is, the greater the number of available frequencies, the higher the ability to follow the input traffic behavior with the most appropriate clock frequency, and consequently the higher the energy saving gain. However, a high number of clock frequencies could cause too frequent switches and therefore QoS degradation. For this reason, the best tradeoff between energy saving and QoS performance can be achieved by using an appropriate set Φ of clock frequencies that is a subset of $\overline{\Phi}$. In addition, we have to take into account that the choice of the particular subset Φ has to depend on the input traffic, that is, its mean value, its variance, and its autocorrelation. In fact, if the input bit rate, due to its first- and second-order statistics, too frequently crosses the value $B_i^{(M)}$ associated with the clock frequency F_i, this clock frequency should not be used.

Once the set of active frequencies Φ is decided, the Router Governor has to work controlling that the QoS requirements are respected. To achieve this goal, indicating the generic ith clock frequency in the set Φ as F_i, we define the Router Governor policy as follows.

Rule 1. If the clock frequency was previously set to F_i (see Figure 2(a), where $i = 3$) and the current input bit rate B_{IN} is greater than $B_i^{(M)}$ ($B_3^{(M)}$ in Figure 2(a)), then the clock frequency is switched to the minimum clock frequency belonging to Φ that does not cause losses (F_4 in Figure 2(a)).

Rule 2. If the clock frequency was previously set to F_i (see Figure 2(b) where $i = 4$) and the current input bit rate B_{IN} is lower than $B_{i-1}^{(M)}$ (e.g., lower than $B_3^{(M)}$ in Figure 2(b)), then it can be switched down to a value F_k less than F_i, but not less than the minimum clock frequency belonging to Φ that does not cause losses (i.e., F_2 in Figure 2(b)). However, since a frequency switch causes a QoS degradation, this is done with a probability $p_G(B_{IN}, i, k)$ which is adaptive to the current input bit rate B_{IN}: the greater the distance between B_{IN} and the maximum bit rate that can be supported by the new clock frequency, the lower the risk of a new frequency switch. To this purpose, referring to the example illustrated in Figure 2(b), the switching probability is defined as follows:

(i) the new clock frequency is set to F_2 with a probability:

$$p_G(B_{IN}, 4, 2) = \delta \frac{B_2^{(M)} - B_{IN}}{B_4^{(M)} - B_{IN}}, \tag{1}$$

(ii) if the result of the previous draw was negative, and so the clock frequency was not set to F_2, the new clock frequency is set to F_3 with a probability:

$$p_G(B_{IN}, 4, 3) = \delta \frac{B_3^{(M)} - B_{IN}}{B_4^{(M)} - B_{IN}}, \tag{2}$$

(iii) if the previous draw is negative again, that is, the clock frequency is not set to F_3, the clock frequency remains F_4.

Generally speaking, if the current clock frequency is F_i and the input bit rate B_{IN} is lower than $B_{i-1}^{(M)}$, the clock frequency can be changed in the set $\{F_j, \ldots, F_i\}$, where F_j is the minimum clock frequency of Φ not causing loss. More specifically, the clock frequency is set to F_k, with $k \in [j, i]$, with a probability:

$$p_G(B_{IN}, i, k) = \left[\prod_{h=j}^{k-1} \left(1 - \delta \frac{B_h^{(M)} - B_{IN}}{B_i^{(M)} - B_{IN}}\right)\right] \cdot \begin{cases} \delta \dfrac{B_k^{(M)} - B_{IN}}{B_i^{(M)} - B_{IN}} & \text{if } k < i \\ 1 & \text{if } k = i. \end{cases} \tag{3}$$

The term $\delta \in [0, 1]$ allows the designer to make clock frequency switches more or less rare. It is easy to argue that its value plays a very important role in the router performance. The design of the clock frequency subset Φ and the parameter δ will be assisted by the analytical model that will be described in Section 3. In order to follow variations of traffic statistics in a long-term time scale, they can be modified runtime according to continuous measurements done by the Router Governor.

3. Markov Model

In this section we define a discrete-time model of the system described so far in order to capture the behavior of the clock frequency process. Since it depends on the input traffic bit rate according to the Router Governor policy, we define the Markov model state as $S^{(\Sigma)}(n) = (S^{(C)}(n), S^{(I)}(n), S^{(S)}(n))$, where

(i) $S^{(C)}(n) \in \mathfrak{F}^{(C)}$ is the clock frequency process at the generic slot n;

(ii) $S^{(I)}(n) \in \mathfrak{F}^{(I)}$ represents the quantized input traffic bit rate at the generic slot n;

(iii) $S^{(S)}(n) \in \mathfrak{F}^{(S)} = \{0, 1\}$ is the indicator variable of a switch at the generic slot n: $S^{(S)}(n) = 1$ if, in the slot n, the router is switching its clock frequency.

The set of states $\mathfrak{F}^{(C)}$ contains the *active frequencies*, that is, all the clock frequencies belonging to the set Φ. The set $\mathfrak{F}^{(I)}$ contains the considered quantized input traffic values.

Let us define the slot duration as the interval between two consecutive observations of the input bit rate; it will be indicated as Δ. In order to define the model time diagram, let us consider two generic states: $\underline{s}_{\Sigma 1} = (s_{C1}, s_{I1}, s_{S1})$ in the slot n and $\underline{s}_{\Sigma 2} = (s_{C2}, s_{I2}, s_{S2})$ in the slot $n + 1$. We assume the following event sequence.

(1) The first action at the beginning of the slot $n + 1$ is the evaluation of the new value of the input traffic bit rate. This value is obtained by sampling the bit rate values and smoothing the obtained sequence with an EWMA filter with a time constant equal to the time slot Δ.

(2) Then, according to the new value of the input traffic bit rate, the Governor decides the clock frequency for the new slot. Let us recall that, as said so far, a clock frequency modification determines that the router enters in the switching interval, during which some performance degradation occurs; all the clock frequency switching slots will be characterized by the state variable $S^{(S)}(n) = 1$. Let $\overline{T}_F$ be the duration of this period.

(3) Then, at the end of the slot $n + 1$, the system state variables are observed.

Now we can define the generic element of the state transition probability matrix as follows:

$$\begin{aligned} Q^{(\Sigma)}_{[s_{\Sigma 1}, s_{\Sigma 2}]} &= \text{Prob}\left\{S^{(\Sigma)}(n+1) = s_{\Sigma 2} \mid S^{(\Sigma)}(n) = s_{\Sigma 1}\right\} \\ &= Q^{(I)}_{[s_{I1}, s_{I2}]} \cdot \eta^{(C)}_{[s_{C1}, s_{C2}]}(s_{I2}) \cdot Q^{(S)}_{[s_{S1}, s_{S2}]}(s_{C1}, s_{C2}), \end{aligned} \tag{4}$$

where

(i) $Q^{(S)}_{[s_{S1}, s_{S2}]}(s_{C1}, s_{C2})$ is the transition probability of the clock frequency switch indicator variable. It is defined as follows:

$$Q^{(S)}_{[s_{S1}, s_{S2}]}(s_{C1}, s_{C2}) = \begin{cases} 1 & \text{if } (s_{C2} \neq s_{C1}, s_{S1} = 0, s_{S2} = 1) \\ 1 & \text{if } (s_{C2} = s_{C1}, s_{S1} = 0, s_{S2} = 0) \\ \dfrac{\Delta}{\overline{T}_F} & \text{if } (s_{S1} = 1, s_{S2} = 0) \\ 1 - \dfrac{\Delta}{\overline{T}_F} & \text{if } (s_{S1} = 1, s_{S2} = 1) \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

where the term $\Delta/\overline{T}_F$ is the probability that the router leaves the switching period. The first two probabilities are set to 1 because they represent the probability of changing the state variable $S^{(S)}(n)$ from 0 to 1 when a clock frequency switch occurs, and the probability of maintaining $S^{(S)}(n)$ equal to 0 when the router works normally.

(ii) $\eta^{(C)}_{[s_{C1}, s_{C2}]}(s_{I2})$ gives the probability of a clock frequency switch depending on the clock frequency switching law used by the Governor to decide the clock frequency according to the input traffic bit rate. It is set to 0 when, according to the clock frequency switching law, it is not possible that the Governor sets the value of s_{C2} when the input traffic value is s_{I2} and the current clock frequency is s_{C1}. Following the Governor policy illustrated in Section 2, it is defined as follows:

$$\eta^{(C)}_{[s_{C1}, s_{C2}]}(s_{I2}) = \begin{cases} 1 & \text{if } s_{I2} > B^{(M)}_{s_{C1}},\ B^{(M)}_{s_{C2}} = s_{I2} \\ 1 & \text{if } s_{I2} = B^{(M)}_{s_{C1}},\ s_{C2} = s_{C1} \\ p_G(s_{I2}, s_{C1}, s_{C2}) & \text{if } s_{I2} < B^{(M)}_{s_{C1}},\ s_{I2} \leq B^{(M)}_{s_{C2}} \leq B^{(M)}_{s_{C1}} \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

The term $p_G(s_{I2}, s_{C1}, s_{C2})$ is the frequency clock switching probability defined as in (3). As said in Section 2, it is adaptive with the current value of the input bit rate;

(iii) $Q^{(I)}$ is the state transition probability matrix for the quantized input traffic. It is an input of the problem, because it characterizes the traffic crossing the router.

Now, from the matrix $Q^{(\Sigma)}$ we can derive the system steady-state probability array $\underline{\pi}^{(\Sigma)}$ by solving the following system:

$$\begin{aligned} \pi^{(\Sigma)} Q^{(\Sigma)} &= \pi^{(\Sigma)}, \\ \pi^{(\Sigma)} \cdot \underline{1}^T &= 1, \end{aligned} \tag{7}$$

where $\underline{1}^T$ is a column array with all the elements equal to one. Its generic element, $\pi^{(\Sigma)}_{[\underline{s}_\Sigma]}$, is the steady-state probability of the state $\underline{s}_\Sigma = (s_C, s_I, s_S)$.

4. Performance Parameter Derivation

Let us now derive the main QoS parameters, with the aim of both evaluating router performance and supporting Router Governor design.

First let us calculate the mean power consumed by the router when the Governor applies the proposed policy:

$$P_{\text{MEAN}} = \sum_{\forall s_C \in \mathfrak{F}^{(C)}} \sum_{\forall s_I \in \mathfrak{F}^{(I)}} \Psi(s_C, s_I) \cdot \sum_{\forall s_S \in \mathfrak{F}^{(S)}} \pi^{(\Sigma)}_{[s_C, s_I, s_S]}, \tag{8}$$

where the term $\Psi(s_C, s_I)$ in (8) is a model input and represents the power consumed when the router is loaded with an input traffic bit rate of s_I and the clock frequency is s_C.

Now let us calculate the QoS parameters that can be degraded during clock frequency switching periods, according to the switching technique applied by the green router. The following three relevant cases will be considered.

(1) If the router remains frozen during the switching period and all the traffic arrived in that period is lost, as, for example, in the green NetFPGA reference router case [18, 19], the QoS parameter to be considered is the probability of loss occurring during the switching periods. It is defined as

$$P_{\text{Loss}} = \lim_{m \to +\infty} \frac{L(m)}{V(m)} = \frac{\overline{L}}{\overline{V}} = \frac{\sum_{s_C \in \mathfrak{F}^{(C)}} \sum_{s_I \in \mathfrak{F}^{(I)}} s_I \pi^{(\Sigma)}_{[s_C, s_I, 1]}}{\sum_{s_\Sigma \in \mathfrak{F}^{(\Sigma)}} s_I \pi^{(\Sigma)}_{[s_\Sigma]}}, \tag{9}$$

where $L(m)$ and $V(m)$ are the cumulative number of lost and arrived bits in m consecutive slots, respectively. The term $\overline{V}$ is the mean value of arrived bits per slot, while the term $\overline{L}$ represents the mean value of bits lost per slot.

(2) If the router remains frozen during the switching period and all the traffic arrived in that period is buffered, the QoS parameter to be considered is the mean delay suffered by the traffic arrived during the switching periods. It can be represented by the mean number of packets that arrive during a switching period:

$$\overline{D} = \frac{\overline{T}_F}{\Delta} \left[\sum_{\forall s_C \in \mathfrak{F}^{(C)}} \sum_{\forall s_I \in \mathfrak{F}^{(I)}} s_I \cdot \pi^{(\Sigma)}_{[s_C, s_I, 1]} \right], \tag{10}$$

where the term in squared brackets represents the mean traffic loading the router during a switching period, while $\overline{T}_F/\Delta$ represents the mean duration of the switching period expressed in slots.

(3) If a clock frequency switch causes a peak of energy consumption [21], the QoS parameter to be considered is the total mean power consumption, $P^{(\text{switch})}_{\text{MEAN}}$, defined as the sum of the mean value of the consumed power not considering the switching events, P_{MEAN}, and the mean power caused by the switches. Indicating the power consumed during a switch period as P_{switch}, and taking into account that a switch lasts for $\overline{T}_F/\Delta$ slots, the overall mean power can be calculated as follows:

$$P^{(\text{switch})}_{\text{MEAN}} = P_{\text{MEAN}} + \frac{P_{\text{switch}}}{\overline{T}_F/\Delta} \sum_{\forall s_C \in \mathfrak{F}^{(C)}} \sum_{\forall s_I \in \mathfrak{F}^{(I)}} \pi^{(\Sigma)}_{[s_C, s_I, 1]}. \tag{11}$$

The term P_{switch} is an input of the problem, while P_{MEAN} has been derived in (8).

Another important parameter that can be derived by the mean consumed power calculated as in (11) is the power saving percentage achieved by using the proposed Governor policy. Depending on whether we consider the power consumed during switches or not, it can be calculated as follows:

$$\rho = \frac{\left(P_{\text{MAX}} - P_{\text{MEAN}}\right)}{P_{\text{MAX}}} \cdot 100\%, \qquad \rho = \frac{\left(P_{\text{MAX}} - P^{(\text{switch})}_{\text{MEAN}}\right)}{P_{\text{MAX}}} \cdot 100\%, \tag{12}$$

where P_{MAX} is the power consumed if no saving policy is applied.

5. Model Application to the Governor Design

In this section we will apply the proposed analytical model to a case study to show how the model can be used in the Router Governor design. More specifically, as discussed so far, the goal is to design the clock frequency subset Φ and the δ probability term to be used in (3). Applying such a switching probability, the greater the value of the δ parameter, the more accurate is the Router Governor in the following input traffic bit rate variations, so obtaining higher power saving, but consequently increasing the loss probability.

The considered case study is constituted by a router like the NetFPGA reference router [22]. In this case the QoS parameter that is degraded by clock frequency switches is the loss probability, as discussed in the first of the cases listed in Section 4. The duration of the switching period depends on the specific implementation of the frequency scaling capability. In this case study we consider a switching period of about 2 μs: during this time interval the board is not able to process packets and this causes packet losses.

The proposed model is used to solve an optimization problem, finding the subset Φ of active clock frequencies and the probability term δ which maximize the power saving gain ρ, subject to the constraint $P_{\text{Loss}} \leq P^{(T)}_{\text{Loss}}$, where $P^{(T)}_{\text{Loss}}$ is the upper bound for the switching loss probability that can be tolerated, hereinafter also called target loss probability.

To this aim we started from a set of measurements achieved for the 2-frequency NetFPGA platform presented by the same authors in a previous work [18, 19], here extended to the following eight clock frequencies: F_1 = 15.625 MHz, F_2 = 31.25 MHz, F_3 = 46.875 MHz, F_4 = 62.5 MHz, F_5 = 78.125 MHz, F_6 = 93.75 MHz, F_7 = 109.375 MHz, and F_8 = 125 MHz. This set of frequencies constitutes the set $\overline{\Phi}$ presented in Section 2.

Table 1: Nonnull elements of the input traffic transition probability matrix.

Inferior pseudodiagonal		Main diagonal		Superior pseudodiagonal	
Pos	Value	Pos	Value	Pos	Value
		(1, 1)	9.9990$e-001$	(1, 2)	1.0000$e-004$
(2, 1)	3.1569$e-005$	(2, 2)	9.9993$e-001$	(2, 3)	3.5098$e-005$
(3, 2)	6.7811$e-006$	(3, 3)	9.9994$e-001$	(3, 4)	4.8774$e-005$
(4, 3)	4.1255$e-005$	(4, 4)	9.9995$e-001$	(4, 5)	6.3636$e-006$
(5, 4)	1.9848$e-005$	(5, 5)	9.9994$e-001$	(5, 6)	3.8975$e-005$
(6, 5)	3.8314$e-005$	(6, 6)	9.9992$e-001$	(6, 7)	3.8609$e-005$
(7, 6)	1.3970$e-005$	(7, 7)	9.9990$e-001$	(7, 8)	8.6030$e-005$
(8, 7)	1.4286$e-004$	(8, 8)	9.9986$e-001$		

As demonstrated in [18], the consumed power can be modeled as follows:

$$\Psi\left(f_C, B_{\mathrm{IN}}\right) = P_C\left(f_C\right) + KP_E\left(f_C\right) + N_I\left(B_{\mathrm{IN}}\right)\cdot E_p\left(f_C\right) + \\ + R_I\left(B_{\mathrm{IN}}\right)\cdot E_r\left(f_C\right) + R_O E_t\left(f_C\right), \tag{13}$$

where f_C is the CPU clock frequency while B_{IN} is the bit rate of the router input traffic. The term $P_C(f_C)$ is the constant baseline power consumption of the NetFPGA card (without any Ethernet ports connected); $P_E(f_C)$ is the power consumed by each Ethernet port (without any traffic flowing); $E_p(f_C)$ is the energy required to process each packet (parsing, routing lookup, etc.); $E_r(f_C)$ is the energy required to receive, process, and store a byte on the ingress Ethernet interface; $E_t(f_C)$ is the energy required to store, process, and send a byte on the egress Ethernet interface; K is the number of Ethernet ports connected (1 to 4); $N_I(B_{\mathrm{IN}})$ is the input traffic bit rate to the NetFPGA card in packets-per-second (pps); $R_I(B_{\mathrm{IN}})$ is the input rate to the NetFPGA card in bytes-per-second; $R_O(B_{\mathrm{IN}})$ is the output rate from the NetFPGA card in bytes-per-second.

Results achieved by applying the power model in (13) to the considered set of eight frequencies are shown in Figure 3. Further measurements on the power consumption relative to the running of the Router Governor procedures have shown that it is negligible with respect to the power consumption of the board. For this reason it has not been considered here.

In order to achieve the input traffic model, we have quantized a traffic trace measured at the ingress of the DIEEI lab router in eight different bit rate levels, ranging from 0.25 Gbit/s to 3.75 Gbit/s with steps of 0.5 Gbit/s. First- and second-order statistics of that trace, in terms of probability density function (pdf) and autocorrelation function (acf), are represented in Figure 4. Then, solving an inverse eigenvalue problem [23, 24], we derived the input traffic Markov model characterized with the same statistical functions as in Figure 4. The traffic model is constituted by the transition probability matrix $Q^{(I)}$ and the bit rate array $\Gamma^{(I)}$. The matrix $Q^{(I)}$ is a tridiagonal matrix whose nonnull elements are listed in Table 1. The bit rate array is $\Gamma^{(I)} = [0.25, 0.75, 1.25, 1.75,$ $2.25, 2.75, 3.25, 3.75]$ Gbit/s. The considered traffic has a mean value of 2.66 Gbit/s and a standard deviation of 0.946 Gbit/s.

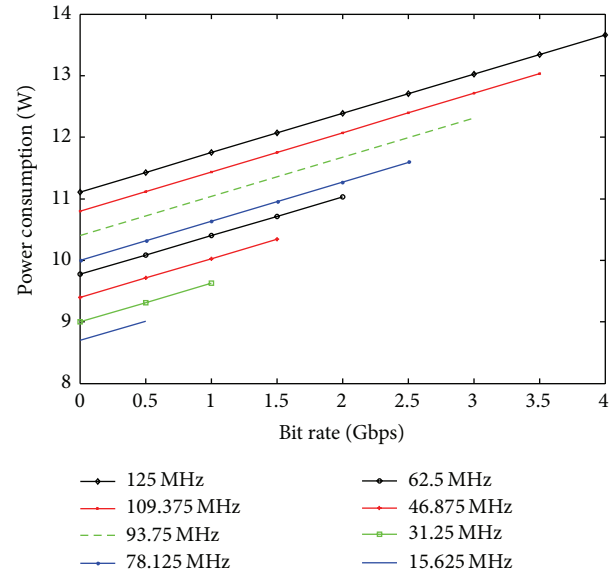
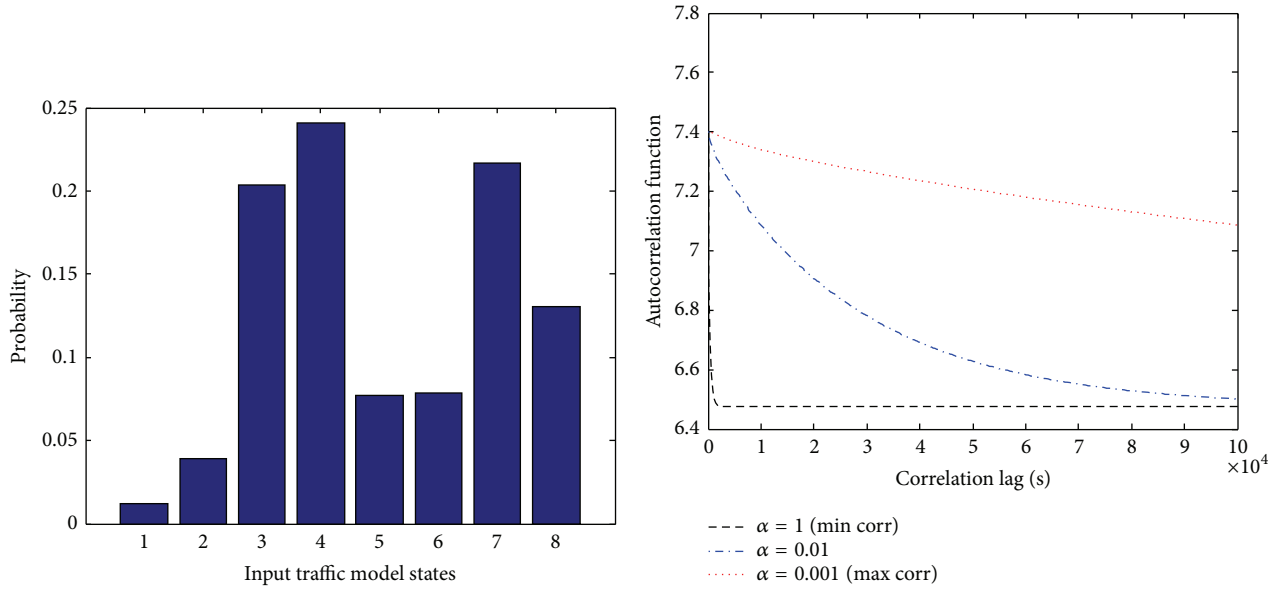


Figure 3: Power consumption model for a router with 8 clock frequencies.

From the traffic model $(Q^{(I)}, \Gamma^{(I)})$ we have derived a set of ten different models, obtained as follows.

(i) $T_i = (Q_i^{(I)}, \Gamma^{(I)})$, with $1 \le i \le 5$, characterized by a transition probability matrix $Q_i^{(I)}$ derived from $Q^{(I)}$ by multiplying the terms of the two pseudodiagonals by a coefficient $\alpha_i \in \{10^4, 10^2, 10^0, 10^{-2}, 10^{-4}, 10^{-6}\}$. The terms of the main diagonals are then calculated such that the sum of each row is equal to one. In this way the traffic modeled by T_1 and T_2 result less correlated than the measured traffic, the traffic modeled by T_3 coincides with the real traffic, while the other models represent more correlated traffic.

(ii) $T_i = (Q_i^{(I)}, \overline{\Gamma}^{(I)})$, with $6 \le i \le 10$, characterized by the same five transition probability matrices of the previous case, that is, $Q_i^{(I)} = Q_{i-5}^{(I)}$, but with a bit rate array $\overline{\Gamma}^{(I)}$ achieved by mirroring the array $\Gamma^{(I)}$ of the



(a) Probability density function

(b) Autocorrelation function

Figure 4: Input traffic first- and second-order statistics.

previous case. By so doing the new pdf is the mirror of the one shown in Figure 4(a), and the new mean value is equal to 1.33 Gbit/s.

Using the analytical system model defined in previous section, we have analyzed the loss probability and the power consumption of the router architecture discussed so far. More in details, we have considered 127 different frequency sets Φ, achieved by choosing from the whole set $\overline{\Phi}$ all the possible subsets containing the highest frequency; that is, $F_8 = 125$ MHz. In other words, the subsets we have considered are $\{F_1, F_8\}, \{F_2, F_8\}, \{F_3, F_8\}, \ldots, \{F_1, F_2, F_8\}, \{F_1, F_3, F_8\}, \{F_1, F_4, F_8\}, \ldots, \{F_1, F_2, F_3, F_4, F_5, F_6, F_7, F_8\}$.

We have solved the optimization problem stated at the beginning of this section, for each of the considered ten traffic models, and versus the target loss probability $P_{\text{Loss}}^{(T)}$. The results are shown in Figure 5, where each point corresponds to the configuration (δ, Φ) that provides the highest power saving for each target loss probability and traffic model.

The reader can notice that, when the value of $P_{\text{Loss}}^{(T)}$ increases, the power saving for all the curves tends to an asymptotic value which mainly depends on the mean value of the input traffic. Therefore this result highlights that the maximum achievable power saving is influenced by the mean value of the input traffic bit rate. Moreover, in the same figure we can also notice that the higher the autocorrelation of the traffic, the higher the power saving for a given target loss probability. It is caused by the fact that, when the traffic autocorrelation is higher, the Router Governor can follow the traffic profile with more rare switches of the clock frequency.

Now, in order to evaluate the impact that the used frequencies have on the router performance (target loss probability and power consumption), we have solved the optimization problem considering a constant number of

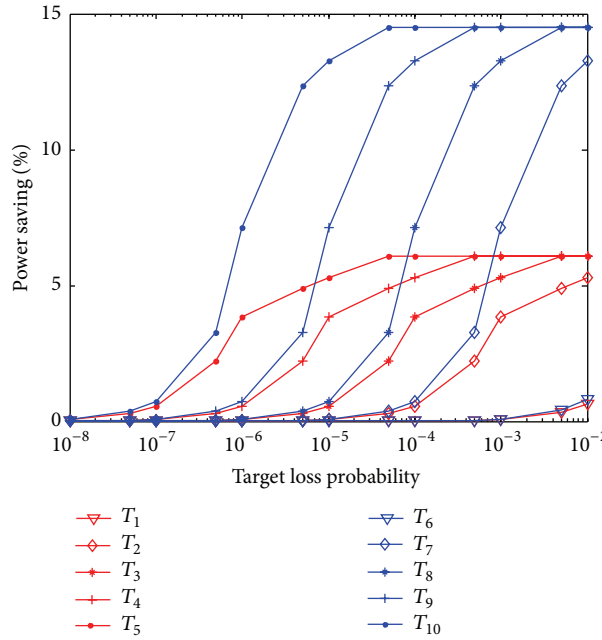
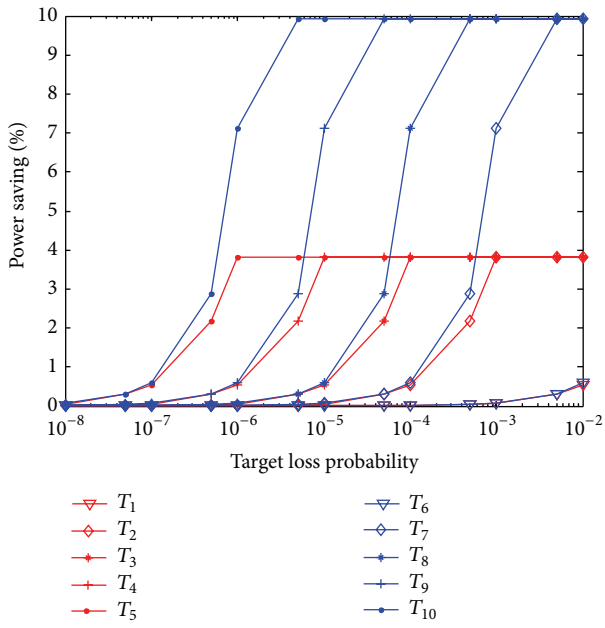
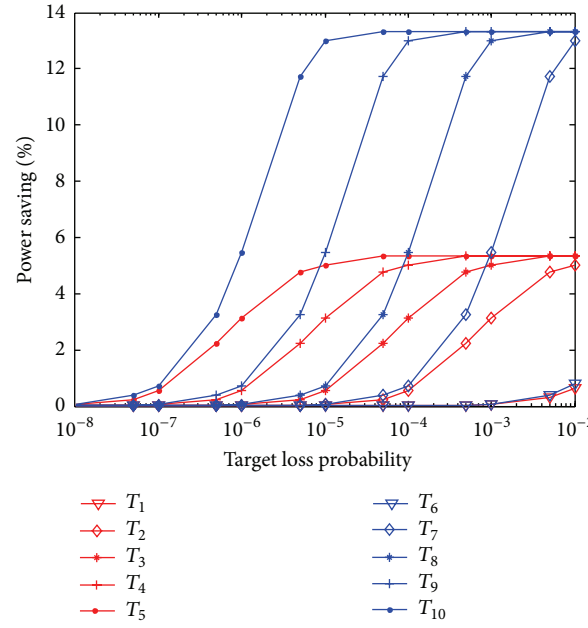


Figure 5: Maximum power saving due to the Router Governor given a maximum loss probability.

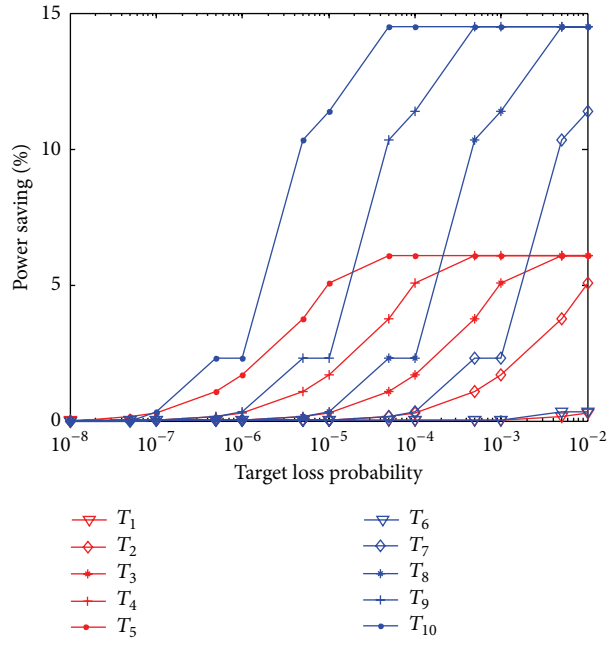
frequencies, leaving the system free to choose the best value of δ and the best set Φ with a number of frequencies equal to the considered one. Figures 6(a), 6(b), and 6(c) show the results for the cases of two, four, and eight frequencies, respectively. We can notice again that the maximum achievable power saving is higher using a higher number of frequencies; this is, because in this case the router processor is able to follow the input traffic more accurately.



(a) Maximum power saving considering only 2 frequencies configurations subset



(b) Maximum power saving considering only 4 frequencies configurations subset



(c) Maximum power saving considering only 8 frequencies configurations subset

FIGURE 6: Power saving versus target loss probability resulted by optimization problem fixing the number of frequencies.

To better investigate the behavior of the Router Governor varying the frequency set and the δ parameter, Figures 7, 8, and 9 show a detailed view of a subset of the cases already represented in Figures 5 and 6. In particular, we consider the cases corresponding to a loss probability target of 10^{-6}. Figure 7 shows the results of the most general optimization problem, solved over the 127 frequency sets described so far. Instead, Figures 8 and 9 present results achieved for the two optimization problems characterized by two and four frequencies, respectively. Such figures explore the frequency configurations and the δ parameter value selected by the optimization algorithm, also showing the power gain of each case. Looking at the above figures we can observe that the Router Governor changes the subset of used frequencies according to both the mean and the autocorrelation of the input traffic.

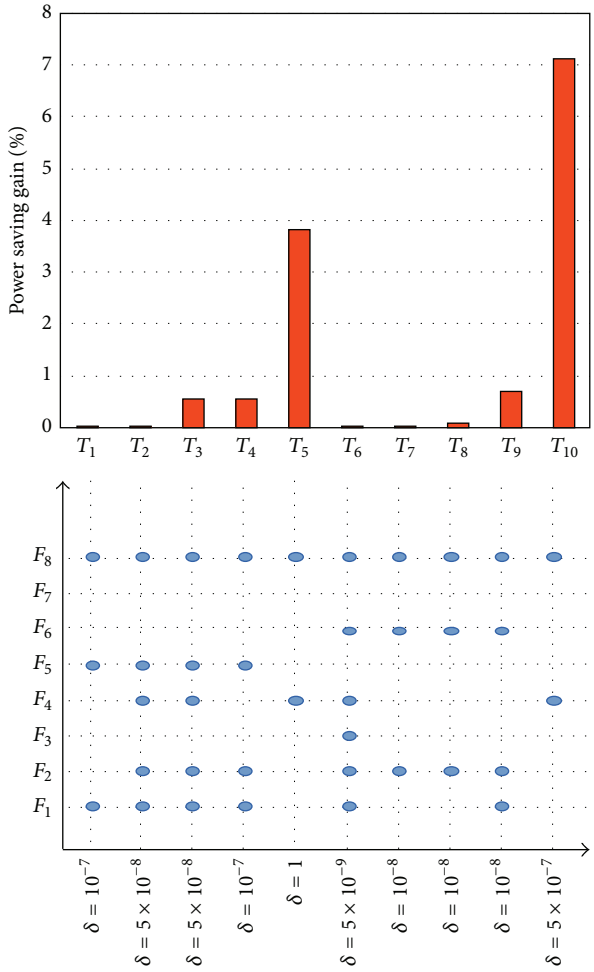


FIGURE 7: Power saving and selected configuration (δ, Φ) corresponding to a target loss probability of 10^{-6}.

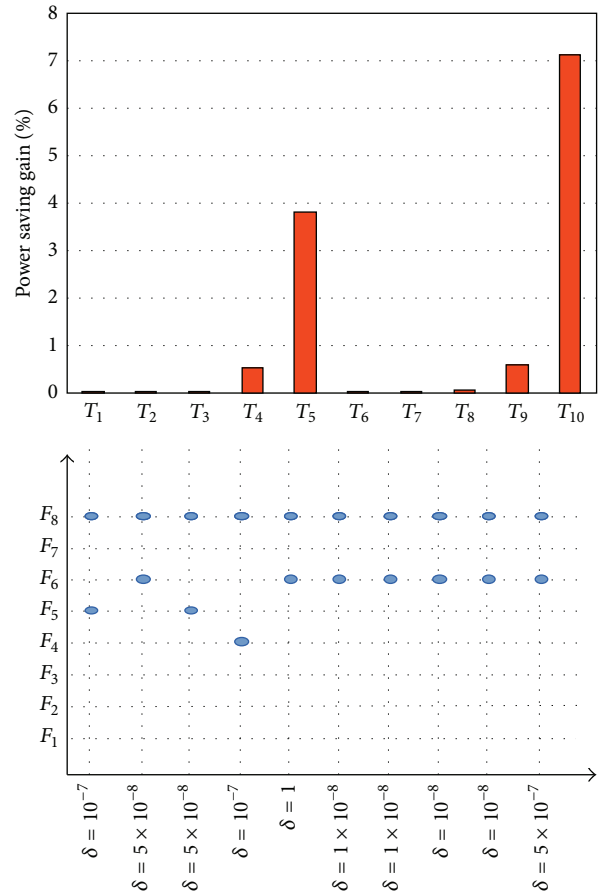


FIGURE 8: Power saving and selected configuration (δ, Φ) corresponding to a target loss probability of 10^{-6}—2-frequency subset.

To better understand the results of the optimization algorithm, we need to further analyze those figures, and doing that, we have to take into account that for cases T_1 and T_6, the traffic is uncorrelated and so the achievable power consumption is very low (see Figures 5 and 6). So, the Router Governor selects a low value of δ to avoid too frequent clock frequency switches. As far as the other cases are concerned, when the traffic autocorrelation increases the number of frequencies can be augmented: in fact, if the Router Governor sets the clock frequency to a value that supports the current input traffic and such frequency remains unchanged for a given amount of time, both the loss probability and the power consumption will be positively influenced.

In Figure 7 the reader can notice that for T_2 and T_3 the optimization problem has selected five frequencies and δ is equal to $5 \cdot 10^{-8}$, whereas for the T_4 case four frequencies have been selected, but the clock frequency is more free to follow the input traffic variations, since δ is equal to 10^{-7}. Instead, in the T_5 case, where the autocorrelation of the input traffic is very high, the algorithm selects only two frequencies but, since $\delta = 1$, leaves the system completely free to change between them every time the input traffic varies.

Regarding Figure 8, same considerations can be formulated, but here we can found a much more evident result for cases T_2, T_3, and T_4: in fact, the higher the autocorrelation of the traffic, the lower the frequencies we can use and therefore the higher the power saving the system can achieve. Also in the same figure, we can notice, for the T_5 case, that the system is free to change the clock frequency following the input traffic (δ is equal to 1). In Figure 9 the optimization algorithm selects four frequencies for each case: for both the cases T_2 and T_3 the δ parameter is equal to $5 \cdot 10^{-8}$, whereas for T_4 and T_5 lower frequencies are selected and the δ parameter leaves the Governor freer to change the clock frequency more often, increasing the power saving and maintaining the same loss probability.

Finally, in order to evaluate the impact of δ on the performance, we have solved the optimization problem for all the 127 sets described so far, but for two given values of δ, that is, 10^{-4} and 10^{-6}. The relative results are in Figures 10(a) and 10(b), respectively. First of all, it is easy to notice that the higher the value of δ, the higher the power saving, since the Router Governor can follow the input traffic more accurately: in fact, we can achieve a higher power saving using a δ equal to 10^{-4} rather than 10^{-6}.

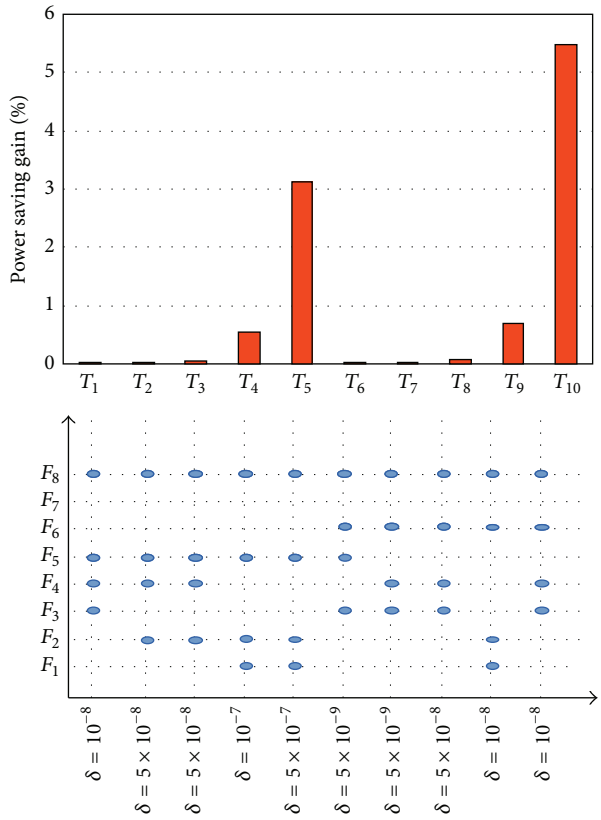
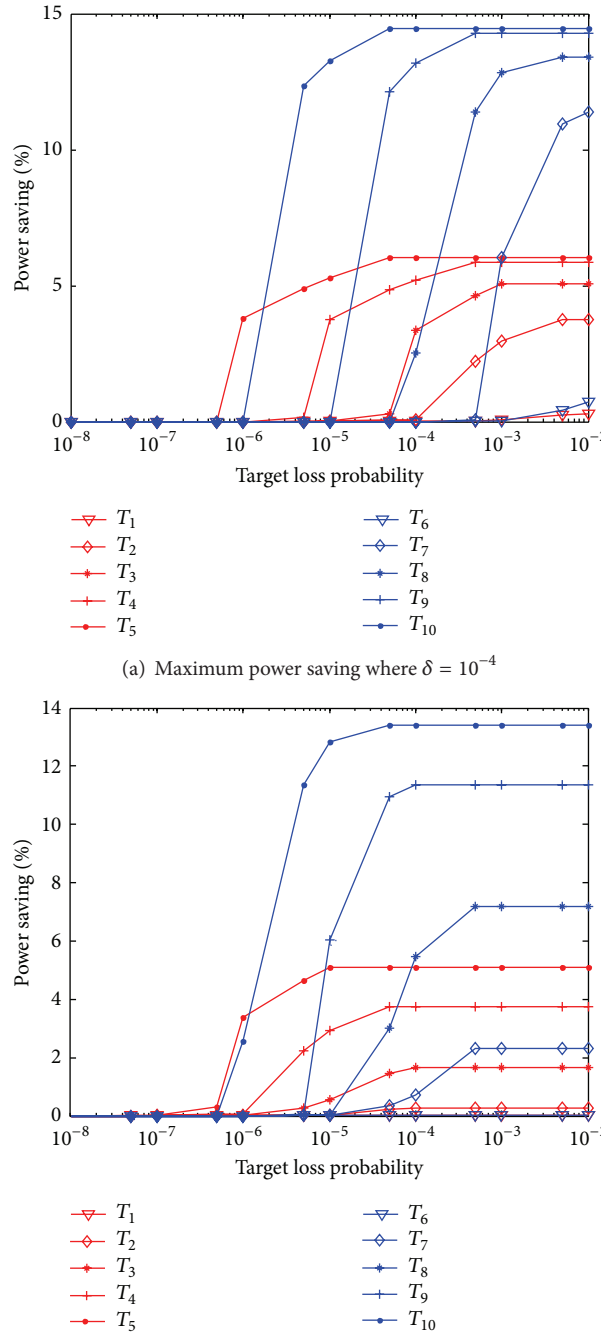


FIGURE 9: Power saving and selected configuration (δ, Φ) corresponding to a target loss probability of 10^{-6}—4-frequency subset.



(a) Maximum power saving where $\delta = 10^{-4}$

(b) Maximum power saving where $\delta = 10^{-6}$

FIGURE 10: Power saving versus target loss probability resulted by optimization problem fixing δ parameter.

It is worth noting that the designed Router Governor can have a strong impact on both power saving and loss probability. In fact, as stated so far, the main important contribution provided by the Governor to the system is to find the best tradeoff between power saving and loss probability. In order to better highlight this matter, in Figures 11 and 12 we have presented power saving and loss probability versus the δ parameter. Reminding that low values of δ lead the system to rarely change the frequency whereas high values of δ lead the system to change the frequency often, accurately following the input traffic. For example, when δ is equal to 1 the system switches the frequency always to the lowest possible one and it corresponds to have high values of power saving, but at the same time high values of loss probability that assumes basically intolerable values (that are very close to 1).

Let us note that all the above figures have been presented to evaluate the impact of the traffic behavior, the parameter δ and the set Φ on the power consumption, and the system performance, but the same figures can also be used by the system designer to choose suitable values of those parameters according to the input traffic, looking for the best tradeoff between power saving and loss probability.

6. Conclusions

In this paper, we have proposed an analytical model to be used to design a Governor for green routers using frequency scaling to save energy. The design aims at limiting the performance worsening due to frequent clock frequency switches. More specifically, the model is used to evaluate the input traffic impacts on the choice of the active router clock

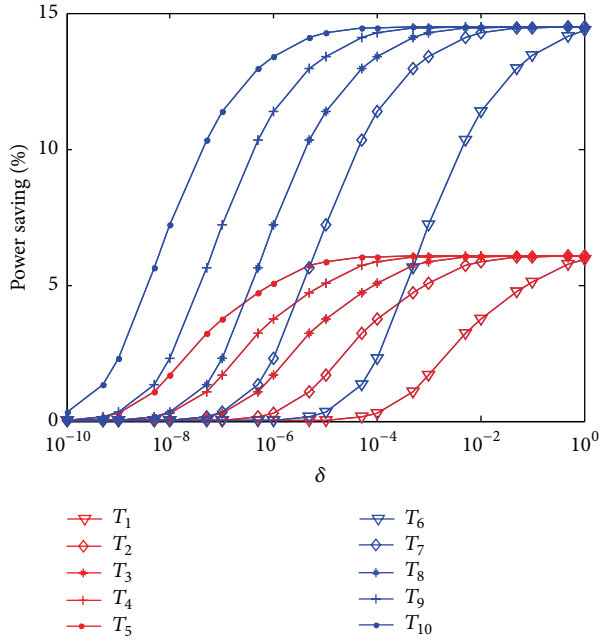


Figure 11: Power saving versus δ parameter.

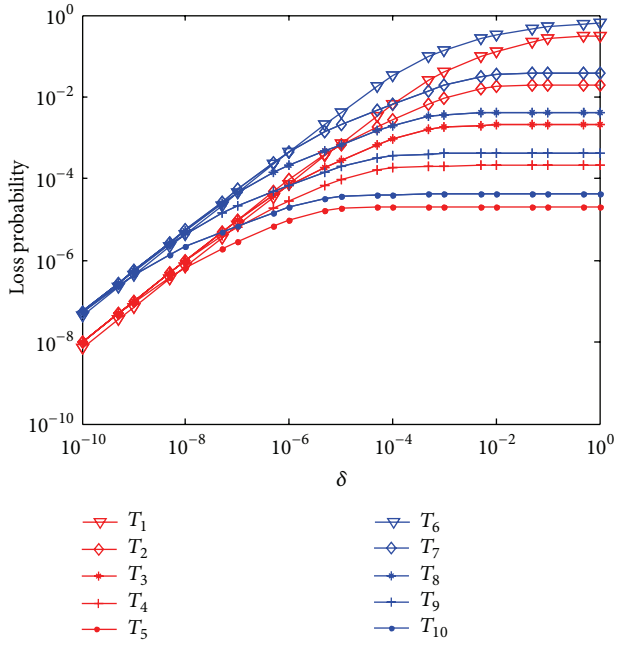


Figure 12: Loss probability versus δ parameter.

frequencies and on the overall green router performance. A case study based on the open NetFPGA reference router is considered to show how the proposed model can be easily applied to a real case scenario.

The model allows the manufacturers to evaluate the power saving gain which is possible to obtain when the proposed Router Governor is used. The future directions that we will pursue are related to an extension of the model to capture the behavior of both input and output queues. In addition, we are working to use the achieved results to design a traffic shaper that is able to modify the autocorrelation of the input traffic to maximize the achieved power saving.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable comments which improved the quality of this paper and clarified many important points to the reader. This work was partially supported by the Econet project, funded by the EU through the FP7 call, and the "Programma Operativo Nazionale "Ricerca & Competitività" 2007–2013" within the project "PON04a2_E—SINERGREEN—RES NOVAE—Smart Energy Master per il governo energetico del territorio."

References

[1] M. Pickavet, W. Vereecken, S. Demeyer et al., "Worldwide energy needs for ICT: the rise of power-aware networking," in *Proceedings of the 2nd International Symposium on Advanced Networks and Telecommunication Systems (ANTS '08)*, pp. 1–3, Mumbai, India, December 2008.

[2] J. Chabarek, J. Sommers, P. Barford, C. Estan, D. Tsiang, and S. Wright, "Power awareness in network design and routing," in *Proceedings of the 27th IEEE Communications Society Conference on Computer Communications (INFOCOM '08)*, Phoenix, Ariz, USA, April 2008.

[3] A. P. Jardosh, G. Iannaccone, K. Papagiannaki, and B. Vinnakota, "Towards an energy-star WLAN infrastructure," in *Proceedings of the 8th IEEE Workshop on Mobile Computing Systems and Applications (HOTMOBILE '07)*, pp. 85–90, Washington, DC, USA, February 2007.

[4] M. Gupta and S. Singh, "Greening of the internet," in *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM '03)*, New York, NY, USA.

[5] S. Nedevschi, L. Popa, G. Iannaccone, S. Ratnasamy, and D. Wetherall, "Reducing network energy consumption via sleeping and rateadaptation," in *Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation*, 2008.

[6] L. Benini, A. Bogliolo, G. A. Paleologo, and G. De Micheli, "Policy optimization for dynamic power management," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 18, no. 6, pp. 813–833, 1999.

[7] T. Šimunić, L. Benini, P. Glynn, and G. De Micheli, "Event-driven power management," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 20, no. 7, pp. 840–857, 2001.

[8] H. Jung and M. Pedram, "Dynamic power management under uncertain information," in *Proceedings of the Design, Automation and Test in Europe Conference and Exhibition (DATE '07)*, pp. 1–6, Nice, France, April 2007.

[9] Q. Qiu, Y. Tan, and Q. Wu, "Stochastic modeling and optimization for robust power management in a partially observable system," in *Proceedings of the Design, Automation and Test in*

Europe Conference and Exhibition (DATE '07), pp. 1–3, Nice, France, April 2007.

[10] R. Bolla, R. Bruschi, F. Davoli, and F. Cucchietti, "Energy efficiency in the future internet: a survey of existing approaches and trends in energy-aware fixed network infrastructures," *IEEE Communications Surveys and Tutorials*, vol. 13, no. 2, pp. 223–244, 2011.

[11] "Econet," 2010, http://www.econet-project.eu/.

[12] "Trend," 2010, http://www.fp7-trend.eu/.

[13] "Greentouch," 2011, http://www.greentouch.org/.

[14] Cisco, "Ciscoenergywise," 2009, http://www.cisco.com/.

[15] C. Hu, C. Wu, W. Xiong, B. Wang, J. Wu, and M. Jiang, "On the design of green reconfigurable router toward energy efficient internet," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 83–87, 2011.

[16] F. Wenliang and T. Song, "A frequency adjustment architecture for energy efficient router," *ACM SIGCOMM Computer Communication Review*, vol. 42, no. 4, pp. 107–108, 2012.

[17] G. Semeraro, G. Magklis, R. Balasubramonian, D. H. Albonesi, S. Dwarkadas, and M. L. Scott, "Energy-efficient processor design using multiple clock domains with dynamic voltage and frequency scaling.," in *Proceedings of the 8th IEEE International Symposium on High-Performance Computer Architecture*, pp. 29–40, February 2002.

[18] A. Lombardo, D. Reforgiato, V. Riccobene, and G. Schembra, "Modeling temperature and dissipation behavior of an open multi-frequency green router," in *Proceedings of the IEEE Online Conference on Green Communications*, September 2012.

[19] A. Lombardo, D. Reforgiato, V. Riccobene, and G. Schembra, "A Markov model to control heat dissipation in open multi-frequency green routers," in *Proceedings of the SustainIT*, Pisa, Italy, October 2012.

[20] G. Gibb, J. W. Lockwood, J. Naous, P. Hartke, and N. McKeown, "NetFPGA—an open platform for teaching how to build gigabit-rate network switches and routers," *IEEE Transactions on Education*, vol. 51, no. 3, pp. 364–369, 2008.

[21] S. Q. Li, S. Park, and D. Arifler, "SMAQ: a measurement-based tool for traffic modeling and queuing analysis. Part I. Design methodologies and software architecture," *IEEE Communications Magazine*, vol. 36, no. 8, pp. 56–65, 1998.

[22] R. Bruschi, A. Lombardo, C. Panarello, F. Podda, G. E. Santagati, and G. Schembra, "Active window management: reducing energy consumption of TCP congestion control," in *Proceedings of the IEEE International Conference on Communications (ICC '13)*, Budapest, Hungary, June 2013, https://github.com/Caustic/netfpga-wiki/wiki/ReferenceRouterWalk.

[23] S. Q. Li, S. Park, and D. Arifler, "SMAQ: a measurement-based tool for traffic modeling and queuing analysis. Part II. Network applications," *IEEE Communications Magazine*, vol. 36, no. 8, pp. 66–77, 1998.

[24] W. Meng, Y. Wang, C. Hu et al., "Greening the internet using multi-frequency scaling scheme," in *Proceedings of the 26th IEEE International Conference on Advanced Information Networking and Applications (AINA '12)*, pp. 928–935, 2012.

Hindawi Publishing Corporation
Journal of Electrical and Computer Engineering
Volume 2014, Article ID 641720, 8 pages
http://dx.doi.org/10.1155/2014/641720

Research Article

Smart Power Management and Delay Reduction for Target Tracking in Wireless Sensor Networks

Juan Feng, Baowang Lian, and Hongwei Zhao

School of Electronic Information Northwestern Polytechnical University, Xi'an, China

Correspondence should be addressed to Juan Feng; fengjuankh@hotmail.com

Received 3 November 2013; Revised 16 January 2014; Accepted 7 February 2014; Published 11 March 2014

Academic Editor: Vincenzo Eramo

Energy efficiency and data transmission delay are critical issues for mobile target tracking wireless sensor networks, in which abundant sensor nodes are deployed to collect the target information from the sensing field. At present, many existing works have been concentrated on extending network lifetime, while less emphasis was placed on both transmission delay reduction and the adaptive sleep of sensor nodes considering the application constraints. In this paper, we propose a smart power management and delay reduction approach for target tracking based on a grid network structure, where sensor nodes adjust their sleep intervals according to the distance between the node and the moving target. Sensor nodes can distributedly decide their sleeping time using the information from their neighbors. Furthermore, we propose a real-time chain to relay the sensed data for transmission delay reduction. The proposed approach allows sensor nodes that are far away from the target to sleep more and make the target information forward to the sink in time. Experimental results verify that, in contrast to adaptive coordinate and local power management protocols, the proposed approach achieves a significant energy saving while maintaining a short transmission delay.

1. Introduction

Wireless sensor networks (WSNs) have emerged as an attractive technology which can gather information by the collective effort of numerous small sizes, low-cost, and wirelessly connected sensor nodes. Due to many attractive characteristics of sensor nodes, WSNs have very extensive applications. One of the important applications is target tracking, such as vehicle tracking and migration behaviour of animals tracking. In such application scenarios, the sensor nodes collectively monitor the path of moving targets in the sensing area. To inform the sink as quickly as possible when the target occurs, the data transmission delay should be short. Since enormous sensor nodes are always deployed in unattended environment, it is very difficult to replace their battery after the deployment. Therefore, balancing the transmission delay and energy consumption to meet the requirement of the application is the most critical issue.

There are some proposals that also consider these issues, such as [1–4], which propose some power management (PM) approaches. The main idea of PM is dynamically getting nodes to sleep to reduce energy consumption since the power consumption in sleep state is usually much smaller than that in the active or idle mode. Thus, the sensor nodes may be put into a sleep mode as long as possible in order to conserve energy. In order to devise a more efficient PM mechanism, the application constraints should be considered. For instance, in target tracking application, a target moves randomly. We do not know the next position of the target. If we blindly turn the sensor node off during each idling time, we will miss some events. In detail, the existing works are not efficient for target tracking because the sleep time is not adjusted in accordance with the position of the target. Most of the time the nodes consume more energy in active or idle mode, but they do not detect any target.

Moreover, the data transmission delay, which is defined as the time between the moment a source sends a packet and the moment a sink receives the packet, is also a critical issue to be considered in target tracking WSNs. If all the sensor nodes adopt an appropriate sleep time according to the

position of the target, each node has the asynchronous wake-up instant. When a node wants to send data, it has to wait until the relay nodes wake up from the sleep mode so that there will be a long data transmission delay. In [5, 6], the authors adopt chain-based network architecture to transmit sensed data. It is an efficient way for data gathering and achieves high energy efficiency since every node just communicates with its closest neighbours. However, it is not fit for the target tracking because it involves all the sensor nodes in the chain for the data gathering so that long data transmission delay is brought, whereas just the nodes around the target need to report their sensed data in the target tracking application. Lots of unnecessary nodes should be moved to reduce the transmission delay.

Based on the these concepts, we propose a smart power management and delay reduction approach (SPM/DR) that considers the application constraints to exploit sleep and idle states. Our main goal is to choose an optimal sleep time for each node so as to make the system adaptive and energy-efficient without degrading the system performance. This paper is based on a grid network structure where a node is selected as grid head (GH) in each grid and we assumed that a sleep node cannot be communicated with or woken up so that the sleep duration has to be determined based on all the available information when the sensor node goes to sleep. We proposed a real-time chain for transmitting sensed data. In a time instant, small parts of GHs in the chain are kept active to relay data and the other GHs are still in sleep to save energy. The information of the target can be transmitted through the chain without delay. In tracking course, only GHs instead of all the sensor nodes send the PM information to their neighbouring GHs, which improves the distributed and coordinated approaches in the existing work. Then each GH weighs the received information according to the distance between itself and its neighbor GH to calculate the sleep time for its grid members (GMs). Therefore, this way decreases the network traffic of PM information transmission and has the better balance between the transmission delay and energy efficiency.

Moreover, in [7], hierarchically coordinated PM is adopted which divides the network into hierarchical layers. Just one layer is in tracking state in one time instance and the others are in sleep state to save energy. This paper improves the approach proposed in [7] from the following aspects: (1) this paper proposed a real-time chain for transmitting sensed data in order to have the better balance between the transmission delay and energy efficiency; (2) for the sleep strategy, this paper adopts the distributed approach instead in which the central sink controls the type of hierarchical layer and sleep strategy in each hierarchical layer, in which the information of moving target needs to be transmitted to the sink in time in order to make decision. Nevertheless, in SPM/DR, each GH decides its sleep interval by the information from its neighbour GHs; (3) SPM/DR approach improves the sleep strategy of GHs by allowing GHs to have a fixed sleep interval with local sleep strategy.

The rest of this paper is organized as follows. Section 2 reviews the related work. Section 3 specifies the energy consumption and management model. In Section 4, SPM/DR approach is described in detail. Then, Section 5 shows the experimental results of SPM/DR compared with the state-of-the-art approaches. Finally, we conclude the paper in Section 6.

2. Related Works

The lifetime of a sensor network depends highly on the power consumption performed at each sensor node. Recently, many target tracking and sensed data gathering approaches have been proposed for WSNs to achieve satisfactory network lifetime and data transmission delay. In [8, 9], the authors design energy efficient communication process at hardware and system levels. In [1, 10], dynamic power management schemes have been proposed to reduce the power consumption.

In [2, 11], the authors introduce some local PM policies, in which nodes reduce the power consumption by selectively shutting down idle components. Each node keeps a timer recording how long no event has been detected and goes to sleep after this timer times out. After a fixed sleep time, the node returns to active state. The authors in [3] use an adaptive learning tree scheme such that the quality of the shutdown control algorithm depends on the knowledge of the user behaviors. However, these policies are intended for general networks. They do not consider the characteristics of target tracking and use the application constraints in a DPM scheme.

For target tracking in WSN, the authors in [8] propose an adaptive coordinate PM policy, which extends the local PM decision to include the timeout values of the neighbours in the network. Once a timeout occurs within a node, it piggybacks a "timeout" message onto its other regularly scheduled messages for wireless broadcast. The receiving nodes strip this timeout message from the packet and forward the packet to subsequent nodes if necessary. In this way, nodes will be aware of the timeout status of their neighbours and can enter into low-power mode if it and each of its neighbours are simultaneously in a timeout state. In [9], a voting PM policy is proposed, in which each node broadcasts periodically a summary of target detection information. Each node collects this summary from its neighbors. If enough of the neighbors vote that there is no event being detected, then the node can enter the low power state. However, due to the dense nodes in WSN, the nodes in an adjacent area have the similar detecting information. If every node broadcasts its detecting information to its neighbours, it will result in more transmission energy consumption and information redundancy. In addition, when nodes make a PM decision, they do not consider different effect of each neighbor. Because of random network distribution, each neighbor has different distance to the current node so that each neighbor has different effect on the current node. In [10], the authors dynamically change the sleep schedule of sensor nodes considering how many hops from the node to the target. However, when a node far from the sink detects a target, it has to wait until the nodes which are far from the target and close to the sink wake up from sleep mode to transmit data. Thus, there will be a long transmission delay for sensed data.

Moreover, in [12], the authors switch off the idle nodes based on the prediction algorithms which predict the next position of the target by the current data from the sensor node to track the target movement. Nevertheless, the prediction algorithms always have high computation complexity and communication energy cost and need to be implemented on the sink after the sink receives all the sensed data.

3. Energy Model

The energy models of data transmitting are similar to that in [5], which is described in (1). $E_{Tx}(k,d)$ and $E_{Rx}(k)$ represent energy consumption of transmitting and receiving k bits data over a distance d

$$E_{Tx}(k,d) = \left(E_{Tx\text{-elec}} + \varepsilon_{\text{amp}} * d^{\alpha}\right) * k,$$
$$E_{Rx}(k) = E_{Rx\text{-elec}} * k, \tag{1}$$

where $E_{Tx\text{-elec}}$ and $E_{Rx\text{-elec}}$ are distance independent terms that take into account overheads of transmitter and receiver electronics. ε_{amp} [Joule/(bit $\cdot$ m^{α})] is a constant which represents the energy needed to transmit one bit to achieve an acceptable signal-to-noise ratio over a distance d, and α is path loss exponent ($2 \le \alpha \le 5$) which depends on the channel characteristics. According to [5] we assume that $E_{Tx\text{-elec}} = E_{Rx\text{-elec}} = E_{\text{elec}}$.

Sensor node has several sleep states because the components of the nodes can be turn on or shut off. In target tracking, if a target will appear within the sensing area of a node, the node should be awake in advance to sense the target and transmit data. For the other nodes, they remain in the sleep state s_k most of the time and switch to active state s_0 at specified time slots to check if there are sensing or relay tasks in the next time instant. Therefore, if sensor nodes cannot change their states in time, the energy will be wasted or the tracking performance is reduced. Now, the problem is how to formulate a policy for sensor node to transfer between these sleep states to maximize the lifetime of the network.

Each sleep state s_k has power consumption P_k. The transition times to it from active state and back are denoted by $\tau_{d,k}$ and $\tau_{u,k}$, respectively [1]. We define the node sleep states as $i > j$, $P_j > P_i$, $\tau_{d,i} > \tau_{d,j}$, and $\tau_{u,i} > \tau_{u,j}$. From that we can see deeper sleep state has less power consumption but incurs a longer latency and a higher energy to awaken. We assume an event is detected by N_i at some time. N_i finishes processing the event at time t_1 and predicts that the next event occurs at time $t_2 = t_1 + t_i + \tau_{u,k}$. At time t_1, N_i decides if it transfers to sleep. So a sleep time threshold $T_{\text{th},k}$ is utilized to avoid losing event

$$T_{\text{th},k} = \tau_{d,k} + \tau_{u,k}. \tag{2}$$

If $(t_2 - t_1) > T_{\text{th},k}$, N_i can go to sleep state s_k at time t_1 and wake up at t_2. Otherwise, when $(t_2 - t_1) \leqslant T_{\text{th},k}$, N_i should not go to the sleep state. So the saving energy from a state transition can be calculated as follows:

$$\begin{aligned} E_{\text{save},k} &= P_0\left(t_i + \tau_{u,k}\right) - \frac{P_0 + P_k}{2}\left(\tau_{d,k} + \tau_{u,k}\right) \\ &\quad - P_k\left(t_i - \tau_{d,k}\right) \\ &= \frac{P_0 - P_k}{2}\left(2t_i - \tau_{d,k} + \tau_{u,k}\right). \end{aligned} \tag{3}$$

The energy saving makes sense when $E_{\text{save},k} > E_c$, where E_c is the additional energy consumption for the sleep states transition. So we can find out the threshold,

$$T_{\text{th},k} = \frac{1}{2}\left(\tau_{d,k} - \tau_{u,k}\right) + \frac{E_c}{P_0 - P_k}. \tag{4}$$

Then, it is clear that the node should get into a sleep state only when its idle period will be long enough. In this paper, the nodes can estimate their idle period more accurately by information and weights from its neighbours.

4. Smart Power Management and Delay Reduction

Our idea is to exploit the long intervals when there is no target in sensing regions and offer the more sleep time to sensor nodes without degrading the system performance. We divide the PM approach into two subapproaches. First, only fringe nodes are kept alert while interior nodes have more sleep time in surveillance stage. Second, each GH adjusts the sleep time for its GMs in light of the information from its neighbour GHs in tracking stage, and the sensed data is sent to the sink through a real-time chain in order to reduce the transmission delay.

4.1. Network Initialization. We consider a static WSN which is composed of one sink and some randomly distributed sensor nodes N_i, $i \in [1, P]$ in a two-dimensional sensing field, where P is the number of the deployed nodes. The sink has an infinite power supply, and it gathers the sensed information from sensor nodes. The distribution of sensor nodes is mutually independent with density $\lambda = P/S_{\text{sensing}}$, where S_{sensing} is the area of the sensing field. We assume each node is aware of its location after deployment (e.g., using some localization techniques). Let $X_i(x_i, y_i)$, $1 \le i \le P$ be the location of node N_i.

Furthermore, the whole sensing field is divided into small equal size grids and in each grid one node which has the most energy is selected as the GH. In the definition of virtual grid, each pair of nodes in neighbour and diagonal grids can communicate directly with each other [1, 11]. Assume that the transmission range of sensor node is R_t. We size each grid as a $\alpha \times \alpha$ square. In order to meet the definition of virtual grid, in any two adjacent grids, the distance between two possible farthest nodes must not be larger than R_t. Therefore, we get

$$(2\alpha)^2 + (2\alpha)^2 \le R_t^2 \quad \text{or} \quad \alpha \le \frac{R_t}{2\sqrt{2}}. \tag{5}$$

Initially, all the sensor nodes are in idle state and they have the same initial energy. One node in each grid is selected as the GH randomly by broadcasting an announcement (carrying its position and grid ID) after waiting for a random time period $R(T_{ran})$, which is a discrete random variable with the uniform distribution in $[0, T_{ran}]$. The node who first broadcasts its GH announcement in one grid will be the GH. After the process, there is one GH in each grid. In addition, each grid has a grid ID, and the nodes in one grid have the same grid ID. A sensor node can calculate its grid ID (u, v) from its location $X_i(x_i, y_i)$ as

$$u = \left[\frac{x_i - x_0}{\alpha}\right], \qquad v = \left[\frac{y_i - y_0}{\alpha}\right], \tag{6}$$

where [ ] is a symbol which stands for the integer part of the number in it. (x_0, y_0) is the location of the network origin, which is a system parameter set in the network initialization stage. For simplicity, we assume that u and v are positive. Through the initial process, each node can learn the information about other nodes in the same grid and maintains the information, for example, the location, and so on.

4.2. A Real-Time Chain for Transmitting Sensed Data. For target tracking application, another important factor to be considered is the detecting delay. Clearly, when any sensor node detects a target, it needs to report to the sink through multihop routing in time. However, increased energy savings generally come with a penalty of increased detecting delay. To resolve this problem, we consider an important characteristic of the tracking scenario in which once a target is detected, the detecting sensor nodes need to continuously send the move information of the target to the sink. As a result, each detecting node has to calculate the routing for data transmission per round, and it has to wait for the next routing node to be in active mode in order to transmit data. That involves an amount of calculations and long delay.

We propose a real-time chain for transmitting the sensed data. When a target appears, the node who detects the target first will broadcast a message "*found_target*" to other nodes including its GH in SS. Then the GH broadcasts the "*found_target*" message to its neighbour GHs to inform them to get ready for detection. The receiving GHs forward the message to the subsequent GHs until all the GHs receive the message. Then a real-time chain from the first detected node to the sink is formed using any routing algorithm (such as some routing protocol in [13]) as Figure 1 shows. The GHs in this transmission chain should keep in active mode to guarantee the short delay for sensed data transmission. Otherwise, the other GHs keep synchronization and use the local PM policy, which keep a timer recording how long no event has been detected and goes to sleep after this timer times out. After a fixed sleep time, they return to active state. If they detect a target, they keep active at the next time instant. Thus, only the nodes around the target and GHs in the real-time chain should keep active and other nodes can sleep adaptively to save energy. Moreover, the sink can estimate the target position by received data. If the moving distance of the target is larger than a threshold T_d, the sink will inform the GHs in the real-time chain and the adjacent GHs to reform a chain. In this way, the sensed data can be sent to the sink with minimum delay. In addition, a lot of energy is saved because computational costs and the changing frequency between sleep and active modes of routing nodes are reduced.

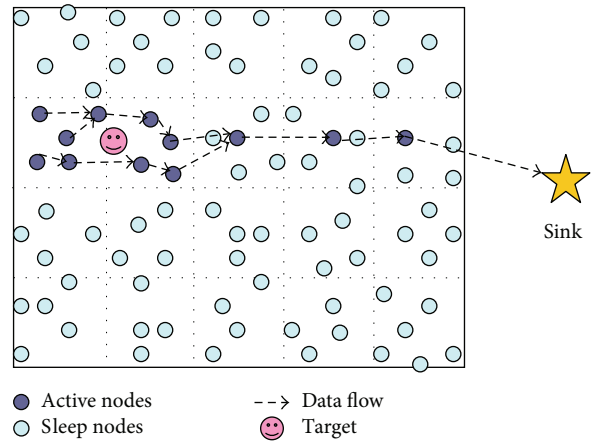


Figure 1: A real-time chain for transmitting sensed data.

4.3. Sleep Policy in Tracking Stage. For tracking stage, the position of the moving target in this time instant has strong correlation with that in next time instant. Therefore, to more effectively make PM decisions and estimate if the target will be in itself sensing area in the next instant, each node can use the results of the motion detection from its neighbors in a coordinate way. As Figure 2 (on the left side) shows, each node broadcast its detected information to its neighbors periodically. When the current node N_c decides if it should go to sleep state, it will use the detected information from its neighbors. However, due to the dense nodes in WSN, the nodes in an adjacent area have the similar detected information. If every node broadcasts its detected information periodically, it will result in more transmission energy consumption and information redundancy. In our approach (on the right side in Figure 2), only the GH who detects a target or receives the target information from its GMs broadcasts the detected information to their neighboring GHs. Each GH can decide if the sleep time of its GMs needs to be changed based on the information it received. If it needs to be changed, the GH recalculates the sleep time for its GMs and informs them in the next active instant, which reduce the energy consumption of the detected information transmission.

4.3.1. One Hop Neighbors Coordination. One hop neighbors coordination means when a GH makes PM decisions, it just considers the detected information from its immediate neighbor GHs. The detected information broadcasted by a GH includes the location of the sender and the distance between the sender and the target. It can be estimated by the strength of the sensing signal. A GH calculates the sleep time for its GMs in three different cases.

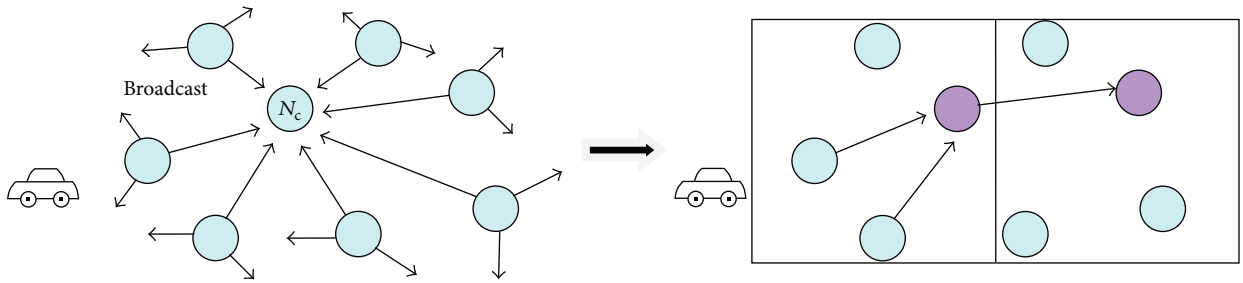


Figure 2: The transmission of the target detecting information.

The first case is that the GH neither detects any target nor receives any target information from others during an active time period. The maximum sleep time is decided by the following equation:

$$\max t_{\text{sleep}}^{\text{tk}} = \frac{\left((\alpha/v_{\max}) - \tau_{d,\text{sp}} - \tau_{u,\text{sp}}\right)}{\eta}, \tag{7}$$

where α is the side length of the grid and η is the success ratio of detecting required by application.

The second case is that the target is detected by the GH or its GMs. The GH keeps active until the target moves out of sensing area of its grid, and it decides the sleep time of its GMs $t_{\text{sleep}}^{\text{tk}} = 0$. Then the GH broadcasts a detected information message to its neighbour GHs so that they can prepare to track the target at the next moment.

The third case is that the GH receives the target information from its neighbors. Thus, the GH can calculate the maximum sleep time for its GMs by the following equation:

$$\max t_{\text{sleep}}^{\text{tk}} = \frac{\left((\min\{d_{\text{GMs},detectNode}\}/v_{\max}) - \tau_{d,\text{sp}} - \tau_{u,\text{sp}}\right)}{\eta}, \tag{8}$$

where $\min\{d_{\text{GMs},detectNode}\}$ is the minimum distance between the GMs of the current grid and the node who detects the target. Similarly, to overlap the active parts of GMs and their GH so that GMs and their GH communicate with each other, we get

$$t_{\text{sleep}}^{\text{tk}}{}' = \left[\frac{\max t_{\text{sleep}}^{\text{tk}}}{T_{\text{GH}}}\right] T_{\text{GH}} - \tau_{u,\text{sp}}, \tag{9}$$

where T_{GH} is the period of GHs. The sleep time of GM is set to the integer multiple of T_{GH} to overlap the active parts of GMs and their GH so that the GMs can send their sensing data to their GHs or receive the PM information from their GHs to adjust their sleep time accordingly. When $t_{\text{sleep}}^{\text{tk}}{}' > T_{\text{th}^1}$ and $t_{\text{sleep}}^{\text{tk}}{}' > T_{\text{th}^2}$ hold, the sleep time of the GMs in tracking stage $t_{\text{sleep}}^{\text{tk}}$ can be set as $\min\{t_{\text{sleep}}^{\text{tk}}{}', t_{\text{sleep}}^{\text{sur}}\}$, in which $t_{\text{sleep}}^{\text{sur}}$ is the sleep time of the GMs in the surveillance stage, otherwise $t_{\text{sleep}}^{\text{tk}} = 0$. When there is change on the sleep time value of GMs, the GH informs its GMs about the updated sleep time. The GMs receive a new sleep time from their GH and adjust the current sleep mode. The sleep state transition of the GM is shown in Figure 3.

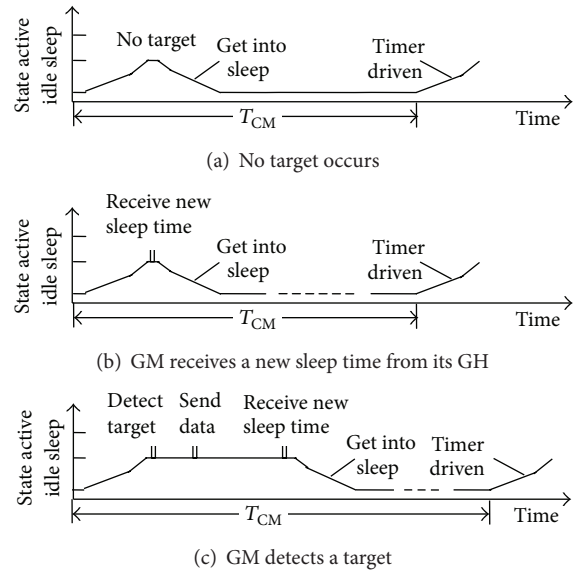


Figure 3: The sleep state transition of the GM.

In this way, sensor nodes will get the target movement information from the GHs in neighboring area and accurately estimate the sleep state and time interval. In addition, transmitting the target information only among GHs instead of that among all the sensor nodes can reduce the communication cost and information redundancy.

4.3.2. Multihop Policy Considerations. In our aforementioned policies, when GHs make the PM decisions, they consider the detected information from the immediate neighbor GHs. We can also increase the coordination to the neighbor GHs located multihop away from the current GH. This requires a balance between the energy consumption of the detected information transmitted to multihops and the energy saving of GMs that can sleep more. We assume that the energy consumption of detected information sending and receiving is E_{br} and E_{re}, respectively. When GHs consider the detected information only from the immediate neighbor GHs, the energy consumption of the detected information transmission is

$$E_{\text{cons}}^{h=1} = E_{\text{br}} + 8E_{\text{re}}, \tag{10}$$

when the detected information is transmitted to the neighbors which are $h = H$ hops far away from the current node

$$E_{\text{cons}}^{h>1} = E_{\text{cons}}^{h=1} + \sum_{h=2}^{H} (8h - 8) E_{\text{br}} + \sum_{h=2}^{H} 8hE_{\text{re}}. \quad (11)$$

When $h = 1$, the energy saving by GMs sleeping is

$$E_{\text{save}}^{h=1} = \left[(P_i - P_s) t_{\text{sleep}}^{tk} + (P_i - P_{\text{tr}^1}) \tau_{d,\text{sp}} + (P_i - P_{\text{tr}^2}) \tau_{u,\text{sp}} \right] \times N_{\text{CMs}}, \quad (12)$$

where N_{CMs} is the number of the GMs in the current grid. Similar to $h = 1$, when $h = H$, the sleep time of GMs is

$$t_{\text{sleep}}^{ml-h'} = \left[\frac{(H\alpha/v_{\max}) - \tau_{d,\text{sp}} - \tau_{u,\text{sp}}}{\eta T_{\text{GH}}} \right] T_{\text{GH}} - \tau_{u,\text{sp}}, \quad (13)$$

and when $h = H$, the energy saving by GMs sleeping is

$$E_{\text{save}}^{h>1} = \left[(P_i - P_s) t_{\text{sleep}}^{ml-h'} + (P_i - P_{\text{tr}^1}) \tau_{d,\text{sp}} + (P_i - P_{\text{tr}^2}) \tau_{u,\text{sp}} \right] \times N_{\text{CMs}}; \quad (14)$$

therefore, when $h = H$, we can obtain the energy saving as follows:

$$E_{\text{save}}^{h>1} - E_{\text{cons}}^{h>1} = (P_i - P_s) N_{\text{CMs}} \left[\frac{(H\alpha/v_{\max}) - \tau_{d,\text{sp}} - \tau_{u,\text{sp}}}{\eta T_{\text{GH}}} \right] T_{\text{GH}} - 4H^2 (E_{\text{br}} + E_{\text{re}}) - 4H (E_{\text{re}} - E_{\text{br}}) + C, \quad (15)$$

where C is a constant which is not related to H,

$$C = N_{\text{CMs}} \left[(P_i - P_{\text{tr}^1}) \tau_{d,\text{sp}} + (P_s - P_{\text{tr}^2}) \tau_{u,\text{sp}} \right] - E_{\text{br}}. \quad (16)$$

To optimize the energy saving, the value of $E_{\text{save}}^{h>1} - E_{\text{cons}}^{h>1}$ needs to be the maximum. Since $P_i - P_s > 0$, $N_{\text{CMs}} > 0$ and $\tau_{d,\text{sp}}$, $\tau_{u,\text{sp}}$, T_{GH}, and η are all the constants. Thus, when the value of y in (17) is maximized, we can get the maximum value of $E_{\text{save}}^{h>1} - E_{\text{cons}}^{h>1}$:

$$y = (P_i - P_s) N_{\text{CMs}} \frac{\alpha}{v_{\max}} H - 4H^2 (E_{\text{br}} + E_{\text{re}}) - 4H (E_{\text{re}} - E_{\text{br}}) + C. \quad (17)$$

To find the maximum y, H is differentiated as

$$\frac{\partial y}{\partial H} = (P_i - P_s) N_{\text{CMs}} \frac{\alpha}{v_{\max}} - 8H (E_{\text{br}} + E_{\text{re}}) - 4 (E_{\text{re}} - E_{\text{br}}); \quad (18)$$

```
(1) while (after initiating or received "req-replace")
(2) do
(3)   node N_i set a timer
(4)   T_i ← CalculateWaitTime (E^i_residual)
(5)   Wait (T_i)
(6)       if wait time expired
(7)         Broadcast ("finish_election")
(8)         select N_i as GH
(9)       end if
(10)      if received ("finish_election" from N_j)
(11)        cancel wait ( )
(12)        select N_j as GH
(13)      end if
(14)  end if
(15) end while
```

Algorithm 1: Selecting GH in one grid.

thus, when $\partial y/\partial H = 0$, we have the maximum y and get H as follows:

$$H = \frac{(P_i - P_s) N_{\text{CMs}} (\alpha/v_{\max}) + 4 (E_{\text{br}} - E_{\text{re}})}{8 (E_{\text{br}} + E_{\text{re}})}; \quad (19)$$

therefore, the maximum saving energy is

$$E_s = \max \left\{ \left(E_{\text{save}}^{h=1} - E_{\text{cons}}^{h=1} \right), \left(E_{\text{save}}^{h>1} - E_{\text{cons}}^{h>1} \right) \right\}. \quad (20)$$

We can select appropriate H to obtain the optimal energy saving policy according to the network parameters. When $t_{\text{sleep}}^{ml-h'} \geq T_{\text{th}^1}$ and $t_{\text{sleep}}^{ml-h'} > T_{\text{th}^2}$ hold, the sleep time of the GMs in tracking stage with multihop GHs coordination can be set as $\min\{t_{\text{sleep}}^{ml-h'}, t_{\text{sleep}}^{\text{sur}}\}$.

4.4. Grid Maintenance. To avoid energy overfull consuming on GH, if the energy of any GH is lower than a threshold T_e, the GH will broadcast a message "*req-replace*" to its GMs. Then the nodes in the same grid reselect the GH. A node with more remaining energy is expected to become a new GH.

The pseudo code of the GH selection is shown in Algorithm 1. When GM receives "*req-replace*" message, it waits a random time to broadcast "*finish_election*" message to compete the GH election. For each GM the residual energy E_{residual}^i is converted to a waiting time T_i (line 4). More residual energy leads to a shorter T_i. Therefore, the GM with the most E_{residual}^i waits for the shortest time before sending "*finish_election*" message to others. Therefore, it can be selected as GH (line 8 and 12). The other nodes stop waiting and give up the GH election as soon as they receive the message. T_i is calculated by the following equation:

$$T_i = T_{\min} + (T_{\max} - T_{\min}) \left(1 - \frac{E_{\text{residual}}^i}{E_{\text{initial}}^i} \right) + R (T_{\text{ran}}), \quad (21)$$

where $T_{\min}$ and $T_{\max}$ are two design parameters, which are used to control the waiting time in a reasonable range.

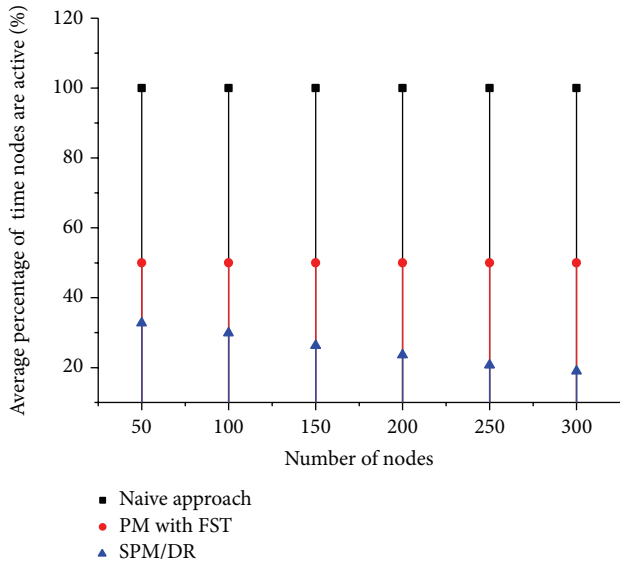


Figure 4: Average percentage of time nodes are active.

E^i_{residual} and E^i_{initial} are the residual and initial energy of node, respectively. E^i_{residual} is divided by E^i_{initial} in order to avoid the case that a node waits too much time when its residual energy becomes very low. Two nodes might have the same residual energy and communication costs and therefore have the same T_i. To avoid this a random time $R(T_{\text{ran}})$ is added. $R(T_{\text{ran}})$ is a discrete random variable with the uniform distribution in $[0, T_{\text{ran}}]$, which is an order of magnitude smaller than $T_{\max}$.

If GH election failed, maybe due to the loss of the broadcast messages, the old GH will continue its role and broadcast the reelection request periodically. For reliability purpose, when a GM fails to send data to its GH for several times (e.g., a GH dies suddenly), it will send a GH reelection request to the GMs.

5. Simulations and Analysis

5.1. Experiment Environment. We assume that there are 400 sensor nodes distributed randomly over an area of 150 m by 150 m. And the sensing range of each sensor node is $r = 15$ m. We also assume each node has an initial energy of 100 J (Joules).

We set $T_e = 1/3$ (initial energy) to avoid energy overfull consuming on GH. T_d is set to 30 m. If the moving distance of the target is larger than T_d, the real-time chain will be adjusted or reformed. The bandwidth of wireless channel is 2.4 Kbps, and the data packet size is 512 bytes.

5.2. Simulation Results. To compare the performance of SPM/DR with other schedules, we implemented the other three approaches, namely, (1) No PM approach (nodes are always on), (2) local PM approach (all the nodes sleep periodically if they did not detect any event in 2 s and the fixed sleep time is also set as 2 s), and (3) adaptive coordinate PM approach.

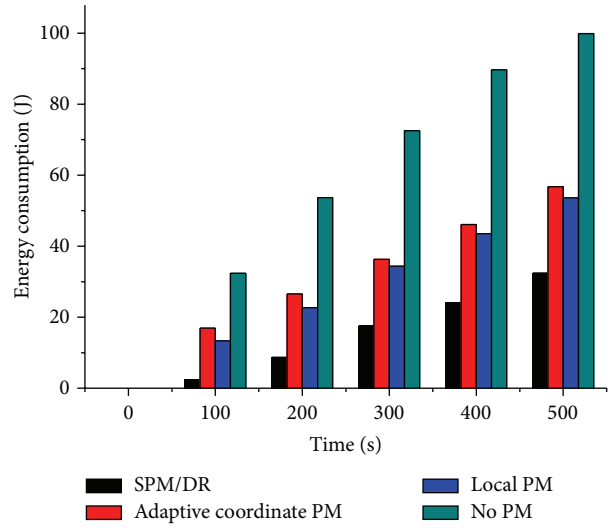


Figure 5: Average energy consumption in surveillance stage.

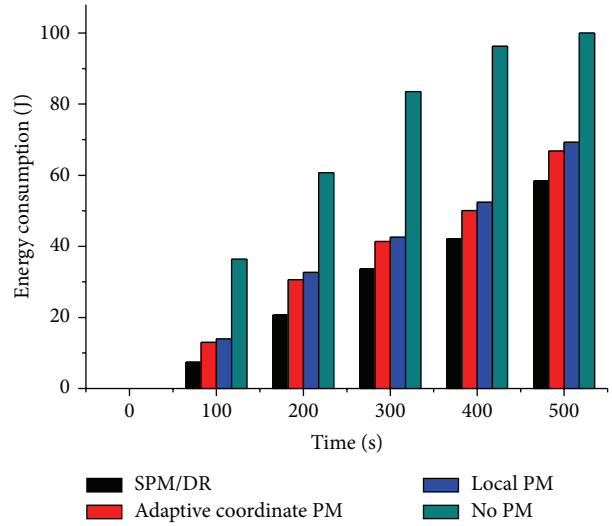


Figure 6: Average energy consumption in tracking stage.

Figure 4 shows the average percentage of time when nodes are in active mode in surveillance stage. It can be seen that most of the nodes in SPM/DR approach are in sleep mode because we just keep the network border nodes alert and the other nodes have a lot of sleep time.

Figure 5 shows the average energy changed with simulation time of the approaches in our simulation. It can be seen that SPM/DR approach can save about 35% more energy compared to adaptive coordinate PM approach in the surveillance stage. No PM approach network consumes the most energy because nodes are always active.

Next, we assume the target enters the field at a random location and moves at a constant speed of 5 m/s. Figure 6 shows the average energy comparison of three approaches. Here, x-axis displays the time of simulation and y-axis displays the average energy consumption of every node at the corresponding time. It can be seen that SPM/DR saves

Table 1: Average transmission delay.

PM approach	Delay
SPM/DR	0.76 ms
Adaptive coordinate PM	0.93 ms
local PM	1.25 ms
No PM	0.59 ms

about 15% more energy compared to adaptive coordinate PM approach.

We can see that adaptive coordinate PM approach consumes more energy than SPM/DR due to more communication cost among neighbor nodes. The local PM approach also has more energy consumption because the nodes far from the target cannot have more sleep time. However our SPM/DR approach consumes the least energy because the sensor nodes in our approach have adaptive sleep time which allowed the nodes far from the target sleep longer.

Table 1 shows the average transmission delay in different approaches. From that we can see no PM approach has the shortest transmission delay since all the nodes are always in active mode. SPM/DR approach performs better than the other two approaches because the GHs in the real-time chain keep active for the sensed data transmission.

In a word, SPM/DR approach can conserve more energy for target tracking WSN. Moreover, it did not degrade tracking performance since we evaluated sleep time adaptively by target move information before the target comes up.

6. Conclusions

This paper proposed a novel smart power management and delay reduction approach (SPM/DR) for tracking target in WSN. It can reduce energy consumption and increase the network lifetime with short data transmission delay. SPM/DR outperforms the other PM approaches by allowing more nodes to have longer sleep intervals and tracks the target by dynamically changing the schedule. Moreover, we use a real-time chain for transmitting the sensed data so that the transmission delay can be reduced. Simulation results proved that SPM/DR performs better than local and adaptive coordinate PM approach for target tracking.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by National Natural Science Foundation of China (61301094) and NPU Foundation for Fundamental Research (NPU-FFR-JCY20130135).

References

[1] A. Sinha and A. Chandrakasan, "Dynamic power management in wireless sensor networks," *IEEE Design and Test of Computers*, vol. 18, no. 2, pp. 62–74, 2001.

[2] P. S. Sausen, M. A. Spohn, and A. Perkusich, "Broadcast routing in wireless sensor networks with dynamic power management and multi-coverage backbones," *Information Sciences*, vol. 180, no. 5, pp. 653–663, 2010.

[3] S. Bhatti and J. Xu, "Survey of target tracking protocols using wireless sensor network," in *Proceedings of the 5th International Conference on Wireless and Mobile Communications (ICWMC '09)*, pp. 110–115, August 2009.

[4] R. Olfati-Saber and P. Jalalkamali, "Collaborative target tracking using distributed Kalman filtering on mobile sensor networks," in *Proceedings of the American Control Conference (ACC '11)*, pp. 1100–1105, San Francisco, Calif, USA.

[5] Q. Mamun, S. Ramakrishnan, and B. Srinivasan, "Selecting member nodes in a chain oriented WSN," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '10)*, April 2010.

[6] Q. Mamun, S. Ramakrishnan, and B. Srinivasan, "An efficient localized chain construction scheme for chain oriented wireless sensor networks," in *Proceedings of the 10th International Symposium on Autonomous Decentralized Systems (ISADS '11)*, pp. 3–9, March 2011.

[7] Juan Feng, Baowang Lian, and Hongwei Zhao, "Hierarchically coordinated power management for target tracking in wireless sensor networks," *International Journal of Advanced Robotic Systems*, vol. 10, article 347, 2013.

[8] N. H. Zamora, J.-C. Kao, and R. Marculescu, "Distributed power-management techniques for wireless network video systems," in *Proceedings of the Design, Automation and Test in Europe Conference and Exhibition*, pp. 564–569, April 2007.

[9] N. H. Zamora and R. Marculescu, "Coordinated distributed power management with video sensor networks: analysis, simulation, and prototyping," in *Proceedings of the 1st ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC '07)*, pp. 4–11, September 2007.

[10] S. Anandamurugan and C. Venkatesh, "Power saving method for target tracking sensor networks to improve the lifetime," *International Journal of Recent Trends in Engineering*, vol. 1, pp. 594–596, 2009.

[11] B. Brock and K. Rajamani, "Dynamic power management for embedded system," in *Proceedings of the IEEE International System-on-Chip (SOC) Conference*, pp. 416–4419, September 2003.

[12] Y. Xu, J. Winter, and W.-C. Lee, "Dual prediction-based reporting for object tracking sensor networks," in *Proceedings of 1st Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services (MOBIQUITOUS '04)*, pp. 154–163, August 2004.

[13] S. K. Singh, M. P. Singh, and D. K. Singh, "Routing protocols in wireless sensor networks—a survey," *International Journal of Computer Science & Engineering Survey*, vol. 1, no. 2, 2010.

Hindawi Publishing Corporation
Journal of Electrical and Computer Engineering
Volume 2014, Article ID 139213, 13 pages
http://dx.doi.org/10.1155/2014/139213

Research Article

Hybrid Optical Switching for Data Center Networks

Matteo Fiorani,[1] Slavisa Aleksic,[2] and Maurizio Casoni[1]

[1] *Department of Engineering "Enzo Ferrari", University of Modena and Reggio Emilia, Via Vignolese 905, 41125 Modena, Italy*
[2] *Institute of Telecommunications, Vienna University of Technology, Favoritenstraße 9-11/E389, 1040 Vienna, Austria*

Correspondence should be addressed to Matteo Fiorani; matteo.fiorani@unimore.it

Received 24 October 2013; Accepted 3 January 2014; Published 27 February 2014

Academic Editor: Vincenzo Eramo

Current data centers networks rely on electronic switching and point-to-point interconnects. When considering future data center requirements, these solutions will raise issues in terms of flexibility, scalability, performance, and energy consumption. For this reason several optical switched interconnects, which make use of optical switches and wavelength division multiplexing (WDM), have been recently proposed. However, the solutions proposed so far suffer from low flexibility and are not able to provide service differentiation. In this paper we introduce a novel data center network based on hybrid optical switching (HOS). HOS combines optical circuit, burst, and packet switching on the same network. In this way different data center applications can be mapped to the optical transport mechanism that best suits their traffic characteristics. Furthermore, the proposed HOS network achieves high transmission efficiency and reduced energy consumption by using two parallel optical switches. We consider the architectures of both a traditional data center network and the proposed HOS network and present a combined analytical and simulation approach for their performance and energy consumption evaluation. We demonstrate that the proposed HOS data center network achieves high performance and flexibility while considerably reducing the energy consumption of current solutions.

1. Introduction

A data center (DC) refers to any large, dedicated cluster of computers that is owned and operated by a single organization. Mainly driven by emerging cloud computing applications data center traffic is showing an exponential increase. It has been estimated that, for every byte of data transmitted over the Internet, 1 GByte is transmitted within or between data centers [1]. Cisco [2] reports that while the amount of traffic crossing the Internet is projected to reach 1.3, zettabytes per year in 2016, the amount of data center traffic has already reached 1.8, zettabytes per year, and by 2016 will nearly quadruple to about 6.6, zettabytes per year. This corresponds to a compound annual growth rate (CAGR) of 31% from 2011 to 2016. The main driver to this growth is cloud computing traffic that is expected to increase sixfold by 2016, becoming nearly two-thirds of total data center traffic. To keep up with these trends, data centers are improving their processing power by adding more servers. Already now large cloud computing data centers owned by online service providers such as Google, Microsoft, and Amazon host tens of thousands of servers in a single facility. With the expected growth in data center traffic, the number of servers per facility is destined to increase posing a significant challenge to the data center interconnection network.

Another issue rising with the increase in the data center traffic is energy consumption. The direct electricity used by data center has shown a rapid increase in the last years. Koomey estimated [3, 4] that the aggregate electricity use for data centers worldwide doubled from 2000 to 2005. The rates of growth slowed significantly from 2005 to 2010, when the electricity used by data centers worldwide showed an increase by about 56%. Still, it has been estimated that data centers accounted for 1.3% of worldwide electricity use in 2010, being one of the major contributors to the worldwide energy consumption of the ICT sector.

The overall energy consumption of a data center can be divided in energy consumption of the IT equipment, energy consumption of the cooling system, and energy consumption of the power supply chain. The ratio between the energy consumption of the IT equipment and the overall energy consumption represents the power efficiency usage (PUE).

The PUE is an important metric that shows how efficiently companies exploit the energy consumed in their data centers. The average PUE among the major data centers worldwide is estimated to be around 1.80 [5], meaning that for each Watt of IT energy 0.8 W is consumed for cooling and power distribution. However, modern data centers show higher efficiency. Google declares that its most efficient data center shows PUE as low as 1.12 [5]. We can then conclude that the major energy savings in modern data centers can be achieved by reducing the power consumption of the IT equipment. The energy consumption of IT equipment can be further divided in energy consumption of the servers, energy consumption of the storage devices, and energy consumption of the interconnection network. According to [6] current data centers networks consume around 23% of the total IT power. When increasing the size of data centers to meet the high requirements of future cloud services and applications, the internal interconnection network will most likely become more complex and power consuming [7]. As a consequence, the design of more energy efficient data center networks is of utmost importance for the scope of building greener data centers.

Current data centers networks rely on electronic switching elements and point-to-point (ptp) interconnects. The electronic switching is realized by commodity switches that are interconnected using either electronic or optical ptp interconnects. Due to the high cross talk and distance dependent attenuation very high data rates over electrical interconnects can be hardly achieved. As a consequence, a large number of copper cables are required to interconnect a high-capacity data center, thereby leading to low scalability and high power consumption. Optical transmission technologies are generally able to provide higher data rates over longer transmission distances than electrical transmission systems, leading to increased scalability and reduced power consumption. Hence, recent high-capacity data centers are increasingly relying on optical ptp interconnection links. According to an IBM study [8] only the replacement of copper-based links with VCSEL-based ptp optical interconnects can reduce the power consumption of a data center network by almost a factor of 6. However, the energy efficiency of ptp optical interconnects is limited by the power hungry electrical-to-optical (E/O) and optical-to-electrical (O/E) conversion required at each node along the network since the switching is performed using electronic packet switching.

When considering future data center requirements, optical switched interconnects that make use of optical switches and wavelength division multiplexing (WDM) technology can be employed to provide high communication bandwidth while reducing significantly the power consumption with respect to ptp solutions. It has been demonstrated in several research papers that solutions based on optical switching can improve both scalability and energy efficiency with respect to ptp interconnects [7, 9, 10]. As a result, several optical switched interconnect architectures for data centers have been recently presented [11–20]. Some of the proposed architectures [11, 12] are based on hybrid switching with packet switching in the electronic domain and circuit switching in the optical domain. The others are based on all-optical switching elements and rely either on optical circuit switching [14, 17] or on optical packet/burst switching [13, 15, 16, 18, 19]. In [20] electronic ToR switches are employed for intrarack communications, while a WDM PON is used for interrack communications. Only a few of these studies evaluate the energy efficiency of the optical interconnection network and make comparison with existing solutions based on electronic switching [12, 17, 20]. Furthermore, only a small fraction of these architectures are proven to be scalable enough to keep up with the expected increase in the size of the data centers [15, 18, 19]. Finally, none of this study addresses the issue of flexibility, that is, the capability of serving efficiently traffic generated by different data centers applications.

With the worldwide diffusion of cloud computing, new data center applications and services with different traffic requirements are continuously rising. As a consequence, future data center networks should be highly flexible in order to serve each application with the required service quality while achieving efficient resource utilization and low energy consumption. To achieve high flexibility, in telecommunication networks hybrid optical switching (HOS) approaches have been recently proposed [21, 22]. HOS combines optical circuit, burst, and packet switching on the same network and maps each application to the optical transport mechanism that best suits its traffic requirements, thus enabling service differentiation directly in the optical layer. Furthermore, HOS envisages the use of two parallel optical switches. A slow and low power consuming optical switch is used to transmit circuits and long bursts, and a fast optical switch is used to transmit packets and short bursts. Consequently, employing energy aware scheduling algorithms, it is possible to dynamically choose the best suited optical switching element while switching off or putting in low power mode the unused ones.

Extending the work presented in [23], in this paper, we propose a novel data center network based on HOS. The HOS switching paradigm ensures a high network flexibility that we have not found in the solutions proposed so far in the technical literature. We evaluate the proposed HOS architecture by analyzing its performance, energy consumption, and scalability. We compare the energy consumption of the proposed HOS network with a traditional network based on optical ptp interconnects. We demonstrate that HOS has potential for satisfying the requirements of future data centers networks, while reducing significantly the energy consumption of current solutions. The rest of the paper is organized as follows. In Section 2 we describe the optical ptp and HOS network architectures. In Section 3 we present the model used for the evaluation of energy consumption. In Section 4 we describe the performed analysis and discuss data center traffic characteristics. In Section 5 we present and discuss the results and, finally, in Section 6 we draw conclusions.

2. Data Centers Networks

2.1. Optical ptp Architecture. Figure 1 shows the architecture of a current data center based on electronic switching and

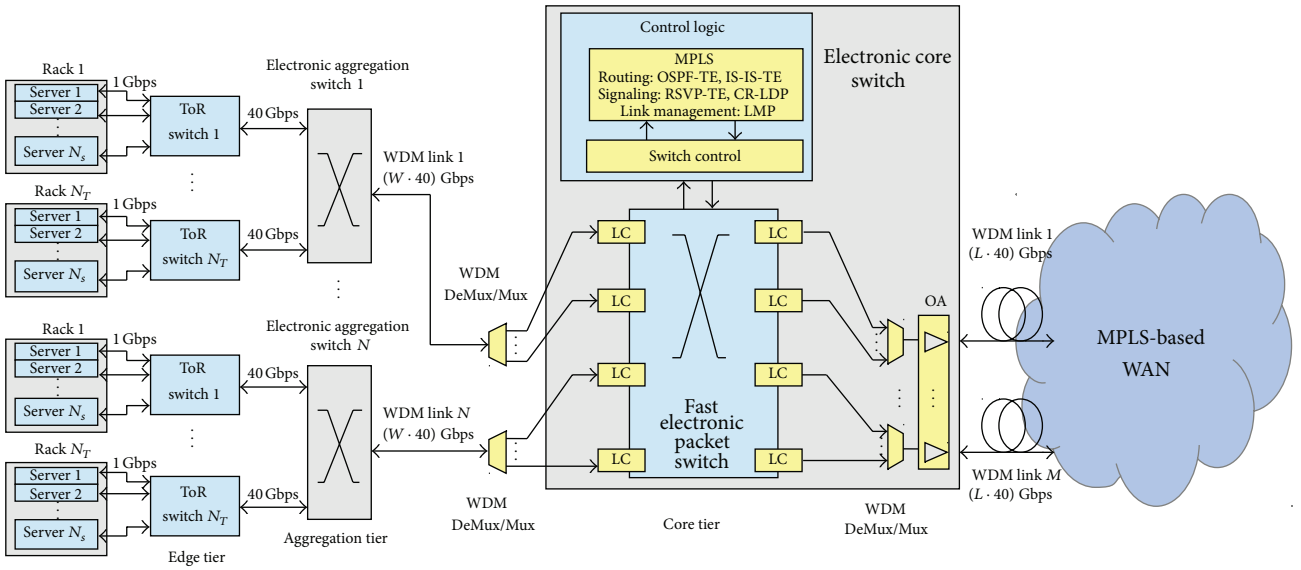


Figure 1: Architecture of a data center employing an optical ptp interconnection network. ToR: top of the rack, OA: optical amplifiers.

optical ptp interconnects. Here, multiple racks hosting the servers are interconnected using a fat-tree 3-Tier network architecture [24]. The 3 tiers of the data center network are the edge tier, the aggregation tier, and the core tier. In the edge tier the top-of-rack (ToR) switches interconnect the servers in the same rack. We assume that each rack contains N_S servers and that each server is connected to a ToR switch through a 1 Gbps link. Although, in future data centers, servers might be connected using higher capacity links, the majority of current data centers still use 1 Gbps links, as reported in [25]. In future works we plan to consider higher capacity per server port and evaluate the effect of increased server capacity. However, it is worth noting that the network performance (e.g., throughput and loss) does not depend on the line data rate, but on the link load, which we consider here as the percentage of the maximum link capacity.

As many as N_T ToR switches are connected to an aggregation switch using 40 Gbps links. The aggregation switches interconnect the ToR switches in the edge tier using a tree topology and are composed of a CMOS electronic switching fabric and electronic line cards (LC), that include power regulators, SRAM/DRAM memories, forwarding engine, and LASER drivers. Each aggregation switch is connected to the electronic core switch through a WDM link composed of W wavelengths channels operated at 40 Gbps. The core switch is equipped with $N \cdot W \cdot 40$ Gbps ports for interconnecting as many as N aggregation switches. Furthermore, the core switch employs $M \cdot L \cdot 40$ Gbps ports for connecting the data center to a wide area network (WAN). We assume that the data center is connected to a WAN employing the MPLS control plane. It is worth noting that the considered optical ptp architecture employs packet switching in all the data center tiers.

The electronic core switch is a large electronic packet switch that comprises three building blocks, namely, control logic, switching fabric, and other optical components. The control logic comprises the MPLS module and the switch control unit. The MPLS module performs routing, signaling, and link management as defined in the MPLS standard. The switch control unit performs scheduling and forwarding functionalities and drives the electronic switching elements. The switching fabric is a single electronic switch interconnecting a large number of electronic LCs. Finally, the other optical components include the WDM demultiplexers/multiplexers (WDM DeMux/Mux) and the optical amplifiers (OA) used as boosters to transmit toward the WAN.

In data centers with many thousands of servers, failures in the interconnection network may lead to losses of a high amount of important data. Therefore, resilience is becoming an increasingly critical requirement for future large-scale data center networks. However, the resilience is out of scope of this study and we do not address it in this paper, leaving it as an open issue for a future work.

2.2. HOS Architecture. The architecture of the proposed HOS optical switched network for data centers is shown in Figure 2. The HOS network is organized in a traditional fat-tree 3-Tier topology, where the aggregation switches and the core switches are replaced by the HOS edge and core node, respectively. The HOS edge nodes are electronic switches used for traffic classification and aggregation. The HOS core node is composed of two parallel large optical switches. The HOS edge node can be realized by adding some minimal hardware modifications to current electronic aggregation switches. Only the electronic core switches should be completely replaced with our HOS core node. As a consequence, our HOS data center network can be easily and rapidly implemented in current data centers representing a good midterm solution toward the deployment of a fully optical data center network. When higher capacities per server, for example, 40 Gbps, will be required, operators can just

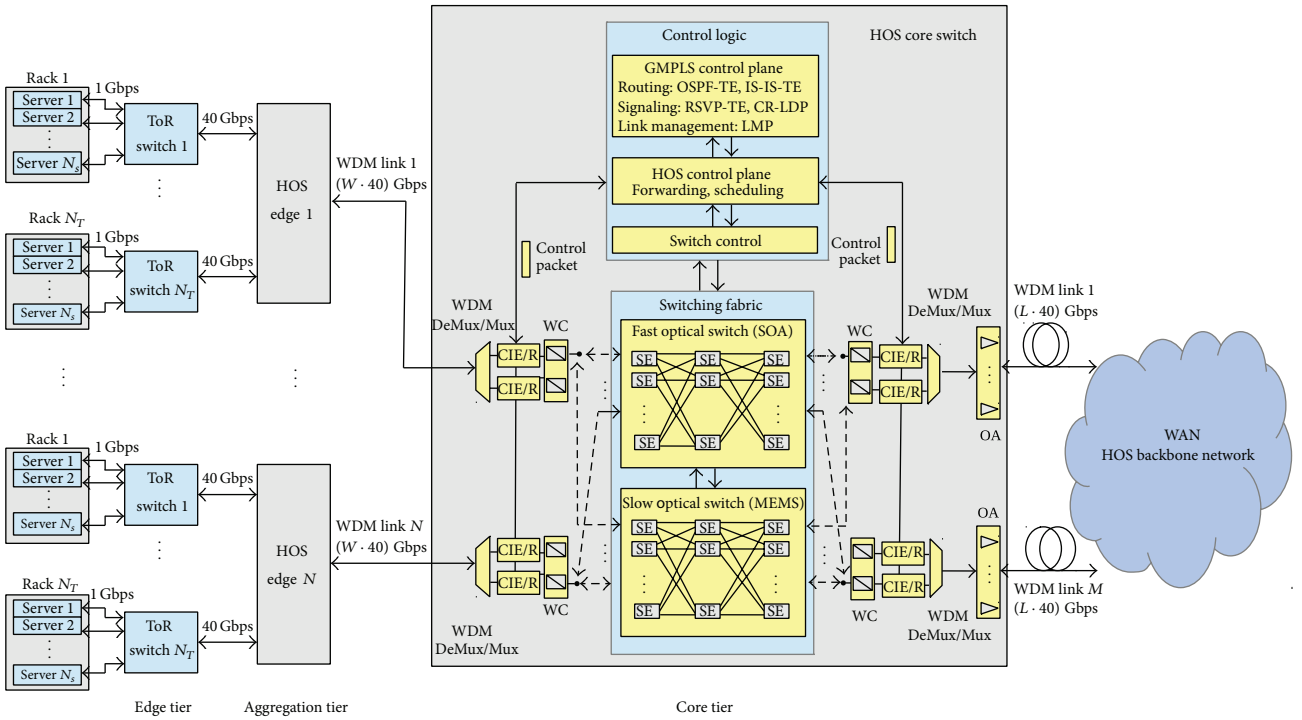


Figure 2: Architecture of a data center employing a HOS interconnection network. CIE/R: control information extraction/reinsertion, WC: wavelength converters.

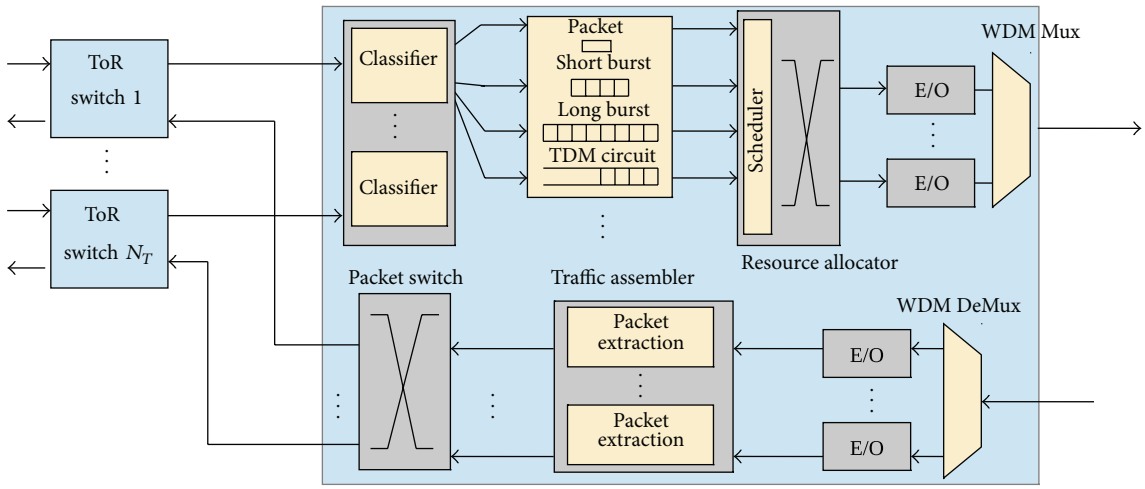


Figure 3: HOS edge node architecture.

connect the servers directly to the HOS edge switches without the need of passing through the electronic ToR switches. In this way it will be possible to avoid the electronic edge tier, meeting the requirements of future data centers and decreasing the total energy consumption. In the long term, it is possible also to think about substituting the electronic HOS edge switches with some optical devices for further increasing the network capacity. This operation will not require any change in the architecture of the HOS core node, which can be easily scaled to support very high capacities. Furthermore, for increased overall performance and energy efficiency we assume that the HOS core node is connected to a HOS WAN [21, 22], but in general the core node could be connected to the Internet using any kind of network technology.

The architecture of a HOS edge node is shown in Figure 3. In the direction toward the core switch the edge node comprises three modules, namely, classifier, traffic assembler, and resource allocator. In the classifier, packets coming from the ToR switches are classified based on their application layer requirements and are associated with the most suited optical transport mechanism. The traffic assembler is equipped with virtual queues for the formation of optical packets,

short bursts, long bursts, and circuits. Finally, the resource allocator schedules the optical data on the output wavelengths according to specific scheduling algorithms that aim at maximizing the bandwidth usage. In the direction toward the ToR switches a HOS edge node comprises packet extractors, for extracting packets from the optical data units, and an electronic switch for transmitting packets to the destination ToR switches.

As for the electronic core switch, we can divide the HOS core node in three building blocks, that is, control logic, switching fabric, and other optical components. The control logic comprises the GMPLS module, the HOS control plane, and the switch control unit. The GMPLS module is used to ensure the interoperability with other core nodes connected to the WAN. The GMPLS module is needed only if the HOS core node is connected to a GMPLS-based WAN, such as the WAN proposed in [21, 22]. The HOS control plane manages the scheduling and transmission of optical circuits, bursts, and packets. Three different scheduling algorithms are employed, one for each different data type, for optimizing the resource utilization, and for minimizing the energy consumption. A unique feature of the proposed HOS control plane is that packets can be inserted into unused TDM slots of circuits with the same destination. This technique introduces several advantages, such as higher resource utilization, lower energy consumption, and lower packet loss probability. For a detailed description of the HOS scheduling algorithms the reader is referred to [26]. Finally, the switch control unit creates the optical paths through the switching fabric. The switching fabric is composed of two optical switches, a slow switch for handling circuits and long bursts, and a fast switch for the transmission of packets and short bursts. The fast optical switch is based on semiconductor optical amplifiers (SOA) and its switching elements are organized in a nonblocking three-stage Clos network. In order to achieve high scalability, 3R regenerators are included after every 9th SOA stages to recover the optical signal, as described in [27]. The slow optical switch is realized using 3D microelectromechanical systems (MEMS). Finally, the other optical components include WDM DeMux/Mux, OAs, tunable wavelength converters (TWCs), and control information extraction/reinsertion (CIE/R) blocks. TWCs can convert the signal over the entire range of wavelengths and are used to solve data contentions.

2.3. HOS Transport Mechanisms. In this section we report a brief description of the HOS concept. For a more detailed explanation regarding the HOS data and control plane, we refer the reader to [22, 23, 26].

The proposed HOS network supports three different optical transport mechanisms, namely, circuits, bursts, and packets. The different transport mechanisms share dynamically the optical resources by making use of a common control packet that is subcarrier multiplexed with the optical data. The use of a common control packet is a unique feature of the proposed HOS network that ensures high flexibility and high resource utilization. Each transport mechanism employs then a particular reservation mechanism, assembly algorithm, and scheduling algorithm according to the information carried in the control packet. For a detailed description of the control plane the reader is referred to [26].

Circuits are long lived optical connections established between the source and destination servers. Circuits are established using a two-way reservation mechanism, with incoming data being queued at the HOS edge node until the reservation has been made through the HOS network. Once the connection has been established data are transmitted transparently toward the destination without any losses or delays other than the propagation delay. In the HOS network circuits are scheduled with the highest priority ensuring a very low circuit establishment failure probability. As a consequence, circuits are well suited for data center's applications with high service requirements and generating long-term point-to-point bulk data transfer, such as virtual machine migration and reliable storage. However, due to relatively long reconfiguration times, optical circuits provide low flexibility and are not suited for applications generating bursty traffic.

Optical burst switching has been widely investigated in telecommunication networks for its potential in providing high flexibility while keeping costs and power consumption bounded. In optical burst switching, before a burst is sent, a control packet is generated and sent toward the destination to make a one-way resource reservation. The burst itself is sent after a fixed delay called offset-time. The offset-time ensures reduced loss probability and enables for the implementation of different service classes. In this paper we distinguish between two types of bursts, namely, short and long bursts, which generate two different service levels. Long bursts are characterized by long offset-times and are transmitted using slow optical switching elements. To generate a long burst incoming data are queued at the HOS edge node until a minimum queue length $L_{\min}$ is reached. After $L_{\min}$ is reached, the burst is assembled using a mixed timer/length approach; that is, the burst is generated as soon as the queue reaches $L_{\max} > L_{\min}$ or a timer expires. The long offset-times ensure long bursts a prioritized handling in comparison to packets and short bursts leading to lower loss probabilities. On the other side, the long offset-times and the long times required for burst assembly lead to large end-to-end delays. Short bursts are characterized by shorter offset-times and are transmitted using fast optical switching elements. To generate a short burst we use a mixed/timer length approach. The short burst is assembled as soon as the queue length reaches a fixed threshold or a timer expires. No minimum burst length is required, as was the case for the long bursts. The shorter offset-times and faster assembly algorithm lead to a higher loss probability and lower delays with respect to long bursts. In [23] we observed that bursts are suited only for delay-insensitive data center applications because of their high latency. Here, we were able to reduce the bursts latency by acting on the thresholds used in the short and long burst assemblers. Still, the bursts present remarkably higher delays than packets and circuits and thus are suited for data-intensive applications that have no stringent requirement in terms of latency, such as MapReduce, Hadoop, and Dryad.

Optical packets are transmitted through the HOS network without any resource reservation in advance. Furthermore, packets are scheduled with the lowest priority. As a consequence they show a higher contention probability with respect to bursts, but on the other hand they also experience lower delays. However, the fact that packets are scheduled with the lowest priority leads to extra buffering delays in the HOS edge nodes, giving place to higher latency with respect to circuits. Optical packets are mapped to data center's applications requiring low latency and generating small and rapidly changing data flows. Examples of data center applications that can be mapped to packets are those based on parallel fast Fourier transform (MPI FFT) computation, such as weather perdition and earth simulation. MPI FFT requires data-intensive all-to-all communication and consequently requires frequent exchange of small data entities.

For a more detailed description of the HOS traffic characteristics we refer the reader to [21, 22].

3. Energy Consumption

We define the power consumption of a data center as the sum of the energy consumed by all of its active elements. In our analysis we consider only the power consumed by the network equipment and thus we exclude the power consumption of the cooling system, the power supply chain, and the servers.

3.1. Optical ptp Architecture. The power consumption of the optical ptp architecture is defined through the following formula:

$$P_{\text{Net}} = N_T \cdot N \cdot P_{\text{ToR}} + N \cdot P_{\text{Aggr}} + P_{\text{Core}}, \quad (1)$$

where P_{ToR} is the power consumption of a ToR switch, P_{Aggr} the power consumption of an aggregation switch, and P_{Core} the power consumption of the core switch. The ToR switches are conventional electronic Ethernet switches. Several large companies, such as HP, Cisco, IBM, and Juniper, offer specialized Ethernet switches for use as ToR switch in data center networks. We estimated the power consumption of a ToR switch by averaging the values found in the data sheets released by these companies. With reference to Figures 1 and 2, without loss of generality, we assume $N_T = W$. As a consequence we can assume that the aggregation switches are symmetric; that is, they have the same number of input and output ports. From now on we will then use N_T to indicate also the number of wavelengths in the WDM links connecting the aggregation and core tiers. The power consumption of an aggregation switch P_{Aggr} is then given by the following formula:

$$P_{\text{Aggr}} = N_T \cdot \left(P_{\text{CMOS}} + P_{\text{LC}}\right). \quad (2)$$

Here, N_T is the number of input/output ports, P_{CMOS} is the power consumption per port of an electronic CMOS-based electronic switch, and P_{LC} is the power consumption of an electronic LC at 40 Gbps.

The power consumption of the electronic core switch is given by the sum of the power consumed by all its building blocks:

$$P_{\text{Core}} = P_{\text{CL}} + P_{\text{SF}} + P_{\text{OC}}, \quad (3)$$

where P_{CL} is the power consumption of the control logic, P_{SF} is the power consumption of the switching fabric, and P_{OC} is the power consumption of the other optical components. P_{CL} includes the power consumption of the MPLS module and the switch control unit. When computing P_{SF} we assume that the electronic ports are always active. This is due to the fact that current electronic switches do not yet support dynamic switching off or putting in low power mode of temporarily unused ports. The reason for that is because the time interval between two successive packets is usually too short to schedule the switching off of the electronic ports. As a consequence, we compute P_{SF} through the following formula:

$$P_{\text{SF}} = \left(N \cdot N_T + M \cdot L\right) \cdot \left(P_{\text{CMOS}} + P_{\text{LC}}\right), \quad (4)$$

where P_{LC} is the power consumption of an electronic LC and P_{CMOS} is again the power consumption per port of an electronic CMOS-based electronic switch. Finally, P_{OC} includes the power consumption of the OAs only, since the WDM DeMux/Mux are passive components. In Table 1 the power consumption of all the elements introduced so far is reported. The values were obtained by collecting and averaging data from a number of commercially available components and modules of conventional switching and routing systems as well as from research papers. The table shows that the main power drainers in a traditional data center network are the electronic LCs, which include the components for packet processing and forwarding [27]. A more detailed explanation on how to compute the power consumption of the electronic core switch is given in [26].

3.2. HOS Architecture. The power consumption of the HOS network architecture is obtained through the following formula:

$$P_{\text{Net}}^{\text{HOS}} = N_T \cdot N \cdot P_{\text{ToR}} + N \cdot P_{\text{Edge}}^{\text{HOS}} + P_{\text{Core}}^{\text{HOS}}, \quad (5)$$

where $P_{\text{Edge}}^{\text{HOS}}$ is the power consumption of the HOS edge node and $P_{\text{Core}}^{\text{HOS}}$ is the power consumption of the HOS core node. The power consumption of the HOS edge node is obtained by summing the power consumption of all the blocks shown in Figure 3:

$$P_{\text{Edge}}^{\text{HOS}} = N_T \cdot \left(P_{\text{Cs}} + P_{\text{As}} + P_{\text{PE}} + P_{\text{CMOS}}\right) + P_{\text{RA}}, \quad (6)$$

where P_{Cs} is the power consumption of the classifier, P_{As} is the power consumption of the traffic assembler, and P_{PE} is the power consumption of a packet extraction module. To compute the power consumption of the classifier and assembler we evaluated the average buffer size that is required for performing correct classification and assembly. We obtained an average required buffer size of 3.080 MByte. The assembler and classifier are realized with two large FPGAs equipped

Table 1: Values of power consumption of the components within the optical ptp and the HOS data center networks.

Components	Power [W]
Top of the rack switch (P_{ToR})	650
Aggregation switch	
Electronic switch (P_{CMOS})	8
Line card (P_{LC})	300
Electronic core switch	
Control logic (P_{CL})	27,096
Optical amplifiers (1 × port)	14
HOS edge node	
Classifier (P_{Cs})	62
Assembler (P_{As})	62
Resource allocator (P_{RA})	296
Packet extractor (P_{PE})	25
HOS core switch	
Control logic ($P_{\text{CL}}^{\text{HOS}}$)	49,638
SOA switch (P_{SOA})	20
MEMS switch (P_{MEMS})	0.1
Tunable wavelength converter (1 × port)	1.69
Control info extraction/reinsertion (1 × port)	17

with external RAM blocks for providing the total required memory size of 3.080 MByte. P_{RA} represents the power consumption of the resource allocator. Again, P_{CMOS} is the power consumption per port of an electronic CMOS-based electronic switch. The power consumption of the HOS core node is obtained by summing the power consumption of the control logic, switching fabric, and other optical components:

$$P_{\text{Core}}^{\text{HOS}} = P_{\text{CL}}^{\text{HOS}} + P_{\text{SF}}^{\text{HOS}} + P_{\text{OC}}^{\text{HOS}}. \quad (7)$$

Here, $P_{\text{CL}}^{\text{HOS}}$ is the sum of the power consumed by the GMPLS module, the HOS control plane, and the switch control unit. When computing $P_{\text{SF}}^{\text{HOS}}$, we assume that the optical ports of the fast and slow switches are switched off when they are inactive. This is possible because when two parallel switches are in use, only one must be active to serve traffic from a particular port at a specified time. In addition, because circuits and bursts are scheduled a priori, the traffic arriving at the HOS core node is more predictable than the traffic arriving at the electronic core switch. We then compute the power consumption of the HOS switching fabric through the following formula:

$$P_{\text{SF}}^{\text{HOS}} = N_{\text{fast}}^{\text{AV}} \cdot P_{\text{SOA}} + N_{\text{Slow}}^{\text{AV}} \cdot P_{\text{MEMS}}. \quad (8)$$

Here, $N_{\text{fast}}^{\text{AV}}$ and $N_{\text{Slow}}^{\text{AV}}$ are, respectively, the average number of active ports of the slow and fast switches obtained through simulations. P_{SOA} and P_{MEMS} are, respectively, the power consumption per port of the SOA-based and MEMS-based switches. The average number of active ports for a specific configuration is obtained through simulations. Finally, $P_{\text{OC}}^{\text{HOS}}$ includes the power consumption of OAs, TWCs, and CIE/R blocks. The values used for the power consumption evaluation of the HOS data center network are included in Table 1. A more detailed explanation on how to compute the power consumption of the HOS core node is given in [26, 27].

4. Modeling Approach

To evaluate the proposed HOS data center network we developed an event-driven C++ simulator. The simulator takes as inputs the parameters of the network and the data center traffic characteristics. The output produced by the simulator includes the network performance and energy consumption.

4.1. Data Center Traffic. In general traffic flowing through data centers can be broadly categorized into three main areas: traffic that remains within the data center, traffic that flows from data center to data center, and traffic that flows from the data center to end users. Cisco [2] claims that the majority of the traffic is the one that resides within the data center accounting for 76% of all data center traffic. This parameter is important when designing the size of the data center and in particular the number of ports of the core node that connects the data center to the WAN. Based on the information provided by Cisco, we designed our data center networks so that the number of ports connecting the core node to the WAN is 24% of the total number of ports of the core node.

In this paper we analyze the data center interconnection network; thus we simulate only the traffic that remains within the data center. To the best of our knowledge a reliable theoretical model for the data center network traffic has not been defined yet. However, there are several research papers that analyze data collected from real data centers [28–30]. Based on the information collected in these papers, the interarrival rate distribution of the packets arriving at the data center network can be modeled with a positive skewed and heavy-tailed distribution. This highlights the difference between the data center environment and the wide area network, where a long-tailed *Poisson* distribution typically offers the best fit with real traffic data. The best fit [30] is obtained with the *lognormal* and *Weibull* distributions that usually represent a good model for data center network traffic. We run simulation using both the lognormal and Weibull distributions. In order to analyze the performance at different network loads, we considered different values for the mean and standard deviation of the lognormal distribution as well as for the shape and scale parameters of the Weibull distribution.

In the considered data center networks, the flows between servers in the same rack are handled by the ToR switches and thus they do not cross the aggregation and core tiers. We define the intrarack traffic ratio (IR) as the ratio between the traffic directed to the same rack and the total generated traffic. According to [28–30], the IR fluctuates between 20% and 80% depending on the data center category and the applications running in the data center. The IR impacts both performance and energy consumption of the HOS network and thus we run simulations with different values for the IR. The IR ratio

has instead a negligible impact on the energy consumption of the optical ptp network. This is due to the fact that in the optical ptp network we do not consider switching off of the core switch ports when being inactive and thus the power consumption is constant with respect to the network traffic characteristics.

In our analysis we set the number of blade servers per rack to 48, that is, $N_S = 48$, that is a typical value used in current high-performance data centers. Although a single rack can generate as much as 48 Gbps, the ToR switches are connected to the HOS edge nodes by 40 Gbps links leading to an oversubscription ratio of 1.2. Oversubscription relies on the fact that very rarely servers transmit at their maximum capacity because very few applications require continuous communication. It is often used in current data center networks to reduce the overall cost of the equipment and simplify data center network design. As a consequence, the aggregation and core tiers of a data center are designed to have a lower capacity with respect to the edge tier.

When simulating the HOS network, we model the traffic generated by the servers so that about 25% of the flows arriving at the edge nodes require the establishment of a circuit, 25% are served using long bursts, 25% are served with short bursts, and the remaining 25% are transmitted using packet switching. We do not consider in this paper the impact of different traffic patterns, that is, the portions of traffic served by circuits, long bursts, short bursts, and packets. In fact, we already evaluated this effect for core networks in [21], where we showed that an increase in traffic being served by circuits leads to slightly higher packet losses and a more evident increase of burst losses. Since in this paper we employ the same scheduling algorithms as in [21], we expect a similar dependence of the performance on the traffic pattern.

4.2. Performance Metrics. In our analysis we evaluate the performance, scalability, and energy consumption of the proposed HOS data center network.

As regards the performance, we evaluate the average data loss rates and the average delays. When computing the average loss rates, we assume that the ToR switches and HOS edge nodes are equipped with electronic buffers with unlimited capacity and thus they do not introduce data losses. As a consequence, losses may happen only in the HOS core node. The HOS core node does not employ buffers to solve data contentions in the time domain but is equipped with TWCs for solving data contentions in the wavelength domain. We consider one TWC per port with full conversion capacity; that is, each TWC is able to convert the signal over the entire range of wavelengths. We define the packet (burst) loss rate as the ratio between the number of dropped packets (bursts) and the total number of packets (bursts) that arrive at the HOS core switch. Similarly, the circuit establishment failure probability is defined as the ratio between the number of negative-acknowledged and the total number of circuit establishment requests that arrive at the HOS core switch. The delay is defined as the time between a data packet is generated by the source server and when it is received by the destination server. We assume that the IR traffic is forwarded by the ToR switches with negligible delay, and thus we analyze only the delay of the traffic between different racks, that is, the traffic that is handled by the HOS edge and core nodes. The delay is given by the sum of the propagation delay and the queuing delay; that is, $D = D_p + D_q$. The propagation delay D_p depends only on the physical distance between the servers. The physical distance between servers in a data center is usually limited to a few hundreds of meters, leading to negligible values for D_p. We then decided to exclude D_p from our analysis and consider $D = D_q$. The queuing delay includes the queuing time at the ToR switch and the delays introduced by the traffic assembler and resource allocator in the HOS edge switch ($D_q = D_{\text{ToR}} + D_{\text{as}} + D_{\text{ra}}$). The HOS optical core switch does not employ buffers and thus does not introduce any queuing delay. We refer to the packet delay as to the average delay of data packets that are transmitted through the HOS core node using packet switching. Similarly, we define the short (long) burst delay as the average delay of data packets that are transmitted through the HOS core node using short (long) burst switching. Finally, the circuit delay is the average delay of data packets that are transmitted through the HOS core node using circuit switching.

As regards the scalability, we analyze our HOS network for different sizes of the data center. In general, data centers can be categorized in three classes: university campus data centers, private enterprise data centers, and cloud computing data centers. While university campus and private enterprise data centers have usually up to a few thousands of servers, cloud computing data centers, operated by large service providers, are equipped with up to tens or even hundred thousands of servers. In this paper we concentrate on large cloud computing data centers. As a consequence, we vary the data center size from a minimum of 25 K servers up to a maximum of 200 K servers.

As regards the energy consumption, we compute the total power consumed by the HOS and the optical ptp networks using the analytical model described in Section 3. To highlight the improvements introduced by our HOS approach, we compare the two architectures in terms of energy efficiency and total greenhouse gas (GHG) emissions. The energy efficiency is expressed in Joule of energy consumed per bit of successfully transmitted data. The GHG emissions are expressed in metric kilotons (kt) of carbon dioxide equivalent (CO_{2e}) generated by the data center networks per year. To compute the GHG emissions, we apply the conversion factor of 0.356 $KgCO_{2e}$ emitted per KWh, which was found in [31].

5. Numerical Results

In this section we show and discuss the results of the performed study. Firstly, we present the data loss rates, secondly, we report the network delays, and, finally, we analyze the energy consumption. We take into consideration several parameters, namely: network load, number of servers, traffic distribution, and IR ratio. We define the load as the ratio between the total amount of traffic offered to the network and the maximum amount of traffic that can be handled by the network. On the other side, the number of servers is given by $N_S^{\text{tot}} = N_S \cdot N_T \cdot N$ and represents the sum of all the servers

Table 2: Reference data center configuration.

Parameter	Value
Number of servers per rack (N_S)	48
Number of ToR per HOS edge node (N_T)	64
Number of HOS edge nodes (N)	32
Number of wavelengths per fiber (W)	64
Number of servers in the data center (N_S^{tot})	98,304
Intrarack traffic ratio (IR)	40%
Traffic distribution	Lognormal
Network load	[65%, 80%]

in the data center. Finally, for the traffic distribution and the IR ratio, we refer to the definitions provided in Section 4.1. The reference data center configuration is reported in Table 2. In the following, we evaluate the response of the network in terms of performance and energy consumption.

5.1. Loss Rates. In this section we show and discuss the average data loss rates in the HOS network.

In Figure 4 we show the average data loss rates in the HOS network as a function of the input load. Two different distributions for the interarrival time of the traffic generated by the servers are considered, that is, lognormal and Weibull. Figure 4 shows that the data loss rates with the lognormal and Weibull distributions present the same trend and very similar values. In the case of the Weibull distribution the loss rates are slightly lower at low and medium loads, but they increase more rapidly with increasing the load. At high loads the loss rates obtained with the Weibull distribution are similar or slightly higher than the loss rates obtained with the lognormal distribution. This effect is particularly evident for the packet loss probability, where the loss rates obtained with the two distributions are more different. Figure 4 also shows that the packet loss rates are always higher than the burst loss rates. This is due to the fact that for packets there is no resource reservation in advance. Due to shorter offset-times, the short bursts show higher loss rates with respect to long bursts, especially for low and moderate loads. Finally, we observe that the circuit establishment failure probability is always null. We conclude that data center applications having stringent requirements in terms of data losses can be mapped onto TDM circuits or long bursts, while applications that are less sensitive to losses can be mapped onto optical packets or short bursts.

In Figure 5 the average data loss rates as a function of the IR are shown. The IR has been varied from 20% to 60%. The figure shows that the higher is the IR and the lower are the data loss rates. This is due to the fact that a higher IR leads to a lower amount of traffic passing through the core switch, thus leading to a lower probability of data contentions. While increasing IR from 20% to 60% the packet and short burst loss rates decrease, respectively, by two and three orders of magnitude. It can also be observed that the difference between the loss rates at 65% and 80% of input load becomes more evident at higher IRs. The circuit established failure probability is always null.

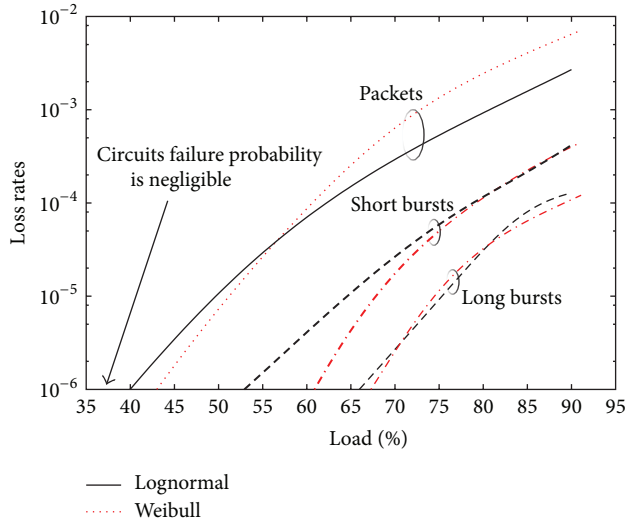


Figure 4: Average data loss rates in the HOS network as a function of the input load. Two different traffic distributions, that is, lognormal and Weibull, are considered.

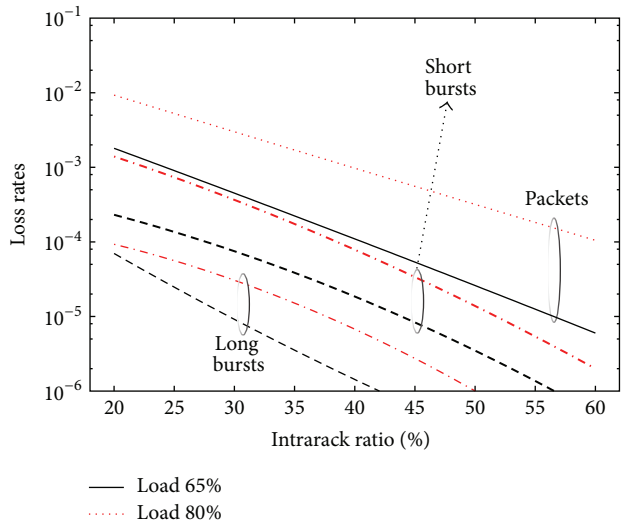


Figure 5: Average data loss rates in the HOS network as a function of the IR at 65% and 80% of offered load.

Finally, in Figure 6 we show the data loss rates as a function of the number of servers in the data center. When changing the size of the data center, we changed both the number of ToR switches per HOS edge node (N_T) and the number of HOS edge nodes (N). We always consider $N_T = W$, in order to have symmetric HOS edge nodes. As a consequence, the higher is N_T and the higher is the number of wavelengths in the WDM links. The smallest configuration was obtained by setting $N = 22$ and $N_T = 24$, achieving a total number of 25,344 servers in the data center. The largest configuration was obtained by setting $N = 50$ and $N_T = 84$ achieving a total number of 201,600 servers. Figure 6 shows that the higher is the size of the data center network and the lower are the loss rates introduced by the HOS core node. This

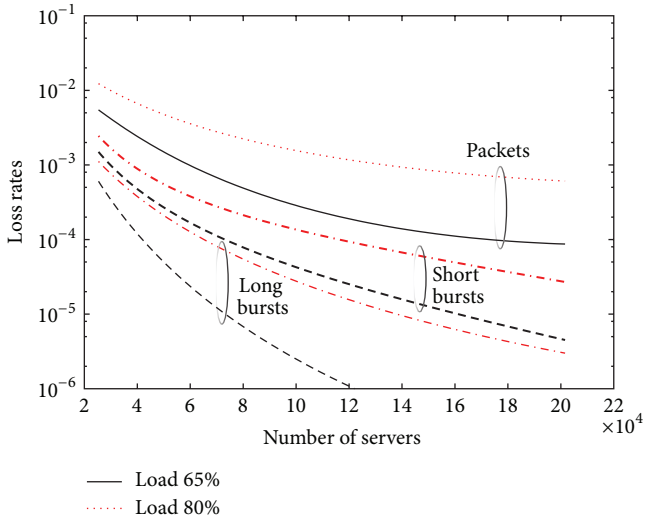


Figure 6: Average data loss rates in the HOS network as a function of the number of servers in the data center. Two different values for the input load, namely, 65% and 80%, are considered.

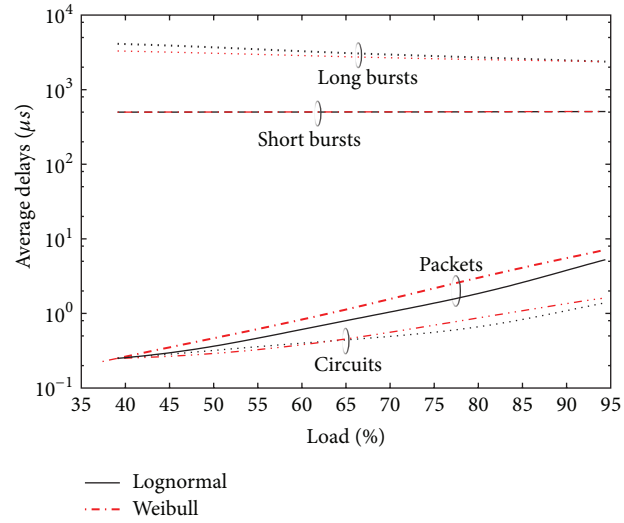


Figure 7: Average delays in the HOS network as a function of the input load and for two different input traffic distributions, namely, lognormal and Weibull.

is due to the fact that in our analysis a higher data center size corresponds to a higher number of wavelengths per WDM link. Since the HOS core node relies on TWCs to solve data contentions, the higher is the number of wavelengths per fiber and the higher is the probability to find an available output resource for the incoming data. This is a unique and very important feature of our HOS data center network, that results in high scalability. In fact, increasing the number of wavelengths per fiber ($N_T = W$) we can scale the size of the data center while achieving an improvement in the network performance. Figure 6 shows that the loss rates, especially the loss rates of the long bursts, decrease by more than one order of magnitude while increasing the number of servers from 25 K to 200 K.

5.2. Delays. In this section we address the network latency. Since there are differences of several orders of magnitude between the delays of the various traffic types, we plotted the curves using a logarithmic scale.

In Figure 7 the average delays as a function of the input load are shown for two different distributions of the interarrival times of packets generated by the servers. The figure shows that the delays obtained with the lognormal and Weibull distributions show the same trends. The largest difference is observed for the delays of packets at high input loads, with the delays obtained with the Weibull distribution being slightly higher. Figure 7 also shows that circuits introduce the lowest delay. To explain this result let us recall the definition of end-to-end delay $D = D_{\text{ToR}} + D_{\text{as}} + D_{\text{ra}}$. For circuits the assembly delay D_{as} is related to the circuit setup delay. Since in our network the circuit setup delay is several orders of magnitude lower than the circuit duration, its effect on the average end-to-end delay is negligible. Furthermore, circuits are scheduled with the highest priority by the resource allocator resulting in negligible D_{ra}. As a consequence, the circuit delay is determined mainly by the delay at the ToR switches D_{ToR}. As can be seen from Figure 7, circuits ensure an average delay below 1.5 μs even for network loads as high as 90%. These values are in line with those presented in [15, 18, 19], where very low-latency optical data center networks are analyzed and are suitable for applications with very strict delay requirements. Packets also do not suffer from any assembly delay, that is, $D_{\text{as}} = 0$, but they are scheduled with low priority in the resource allocator resulting in nonnegligible values for D_{ra}. However, it can be observed that the packet delay remains below 1 μs up to 65% of input load. For loads higher than 65% the packet delays grow exponentially, but they remain bounded to a few tens of μs even for loads as high as 90%. These values are similar to those presented for other optical packet switched architectures, for example, [20], and are suitable for the majority of today's delay-sensitive data center applications.

Short and long bursts are characterized by very high traffic assembler delays D_{as}, which are given by the sum of the time required for the burst assembly and the offset-time. The traffic assembler delay is orders of magnitude higher than D_{ToR} and D_{ra} and thus the end-to-end delay can be approximated with D_{as}. In order to reduce the bursts delays obtained in [23] we acted on the timers and the length thresholds of the burst assemblers. We optimized the short and long burst assemblers and strongly reduced the bursts delays. Still, short and long bursts delays are, respectively, one and two order of magnitude higher than packets delays, making bursts suitable only for delay-insensitive data center applications. Figure 7 shows that short bursts present an almost constant delay attested around 500 μs. Instead, the long burst delay decreases while increasing the input load. This is due to the fact that the higher is the rate of the traffic arriving at the HOS edge node and the shorter is the time required for reaching the long burst threshold L_{min} and starting the process for

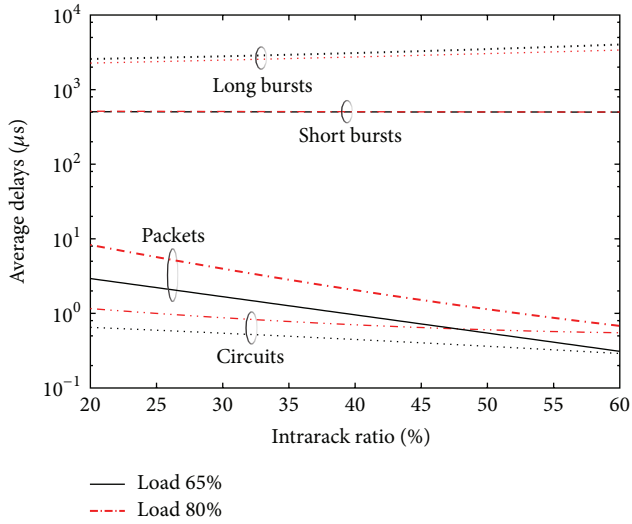


Figure 8: Average delays in the HOS network as a function of the IR at 65% and 80% of offered load.

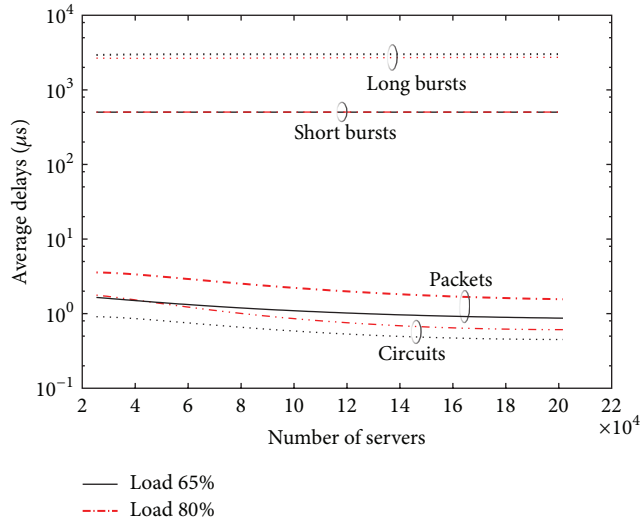


Figure 9: Average delays in the HOS network as a function of the number of servers in the data center. Two different values for the input load, namely, 65% and 80%, are considered.

generating the burst. The minimum long burst delay, which is obtained for very high input loads, is around 2 ms. This delay is quite high for the majority of current data center applications and raises the question if it is advisable or not to use long bursts in future data center interconnects. On the one hand long bursts have the advantage of introducing low loss rates, especially at low and moderate loads, and reducing the total power consumption, since they are forwarded using slow and low power consuming switching elements. On the other hand, it may happen that a data center provider does not have any suitable application to map on long bursts due to their high latency. If this is the case, the provider could simply switch off the long burst mode and run the data center using only packets, short bursts, and circuits. This highlights the flexibility of our HOS approach, that is, the capability of the HOS network to adapt to the actual traffic characteristics.

In Figure 8 we show the average delays in the HOS network as a function of the IR. The figure shows that the circuits and packets delay decrease while increasing the the IR traffic. This is due to the fact that the higher is IR and the lower is the traffic that crosses the ToR switches and the HOS edge nodes in the direction toward the HOS core node. This leads in turn to lower D_{ToR} and lower D_{ra}. In particular, when IR is as high as 60% the D_{ra} for packets becomes almost negligible and the packets delays become almost equal to the circuits delays. As for the long bursts, the higher is IR and the higher are the delays. In fact, a higher IR leads to a lower arrival rate at the HOS edge nodes and, consequently, to a longer assembly delay D_{as}. Finally, the short burst delay is almost constant with respect to IR.

In Figure 9 we show the average delays as a function of the number of servers in the data center. The figure shows that increasing the size of the HOS data center leads to a slight decrease of the end-to-end delays. To explain this fact it is worth remembering that when increasing the number of servers we also increase the number of wavelengths per fiber $W = N_T$ in the WDM links. The higher is N_T and the lower is

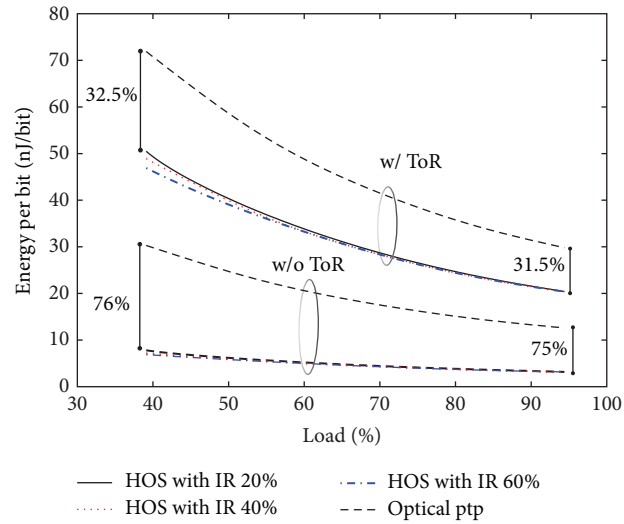


Figure 10: Energy consumption per bit of successfully transmitted data for the HOS and optical ptp networks. Both the cases with and without the ToR switches are shown.

the time required by the resource allocator to find an available output resource where to schedule the incoming data; that is, the higher is N_T and the lower is D_{ra}. This fact again underlines the scalability of the proposed HOS solution.

5.3. Energy Consumption. In this section we present and compare the energy efficiency and GHG emissions of the HOS and the optical ptp data center networks.

In Figure 10 the energy consumption per bit of successfully delivered data is shown as a function of the input load. In the case of the HOS network we consider three different values for IR, namely, 20%, 40%, and 60%. The energy consumption of the optical ptp network is independent

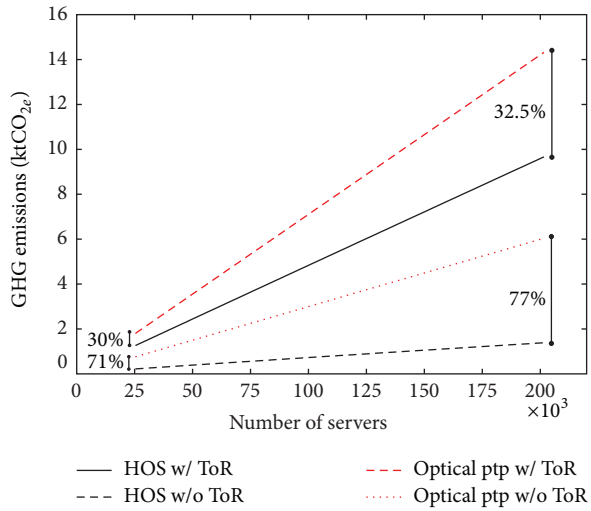


Figure 11: Greenhouse gas emissions per year of the HOS and the optical ptp networks as a function of the size of the data center.

with respect to the IR. Firstly, we consider the overall energy consumption of the data center network and thus we include in our analysis the power consumption of the ToR switches. The electronic ToR switches are the major contributor to energy consumption especially for the HOS network where they consume more than 80% of the total. In the optical ptp network ToR switches are responsible for around 50% of the total energy consumption. Figure 10 shows that the proposed HOS network provides energy savings in the range between 31.5% and 32.5%. The energy savings are due to the optical switching fabric of the HOS core node that consumes considerably less energy with respect to the electronic switching fabric of the electronic core switch. Furthermore, the HOS optical core node is able to adapt its power consumption to the current network usage by switching off temporarily unused ports. This leads to additional energy savings especially at low and moderate loads when many ports of the switch are not used. However, the improvement in energy efficiency provided by HOS is limited by the high power consumption of the electronic ToR switches. In order to evaluate the relative improvement in energy efficiency provided by the use of HOS edge and core switches instead of traditional aggregate and core switches, we show in Figure 10 also the energy efficiency obtained without the energy consumption of the ToR switches. It can be seen that the relative gain offered by HOS is between 75% and 76%. The electronic ToR switches limit then by more than two times the potential of HOS in reducing the data center power consumption, raising the issue for a more energy efficient ToR switch design. Finally, Figure 10 shows that the energy efficiency of the HOS network depends only marginally on the IR traffic ratio. While increasing the IR ratio the energy consumption decreases because a higher IR ratio leads to a lower amount of traffic crossing the HOS core node. Due to the possibility of switching off unused ports, the lower is the amount of traffic crossing the HOS core node and the lower is its energy consumption.

Figure 11 shows the GHG emissions per year of the HOS and the optical ptp networks versus the number of servers in the data center. Again we show both the cases with and without the ToR switches. The figure illustrates that the GHG emissions increase linearly with the number of servers in the data center. In both the cases with and without the ToR switches the GHG emissions of the HOS architecture are significantly lower than the GHG emissions of the optical ptp architecture. In addition, the slopes of the GHG emission curves of the HOS network are lower. In fact, while increasing the number of servers from 25 K to 200 K the reduction in GHG emissions offered by the HOS network increases from 30% to 32.5% when the power consumption of the ToR switches is included and from 71% to 77% when the power consumption of the ToR switches is not included. This is due to the fact that the power consumption of all the electronic equipment depends linearly on the size, while the power consumption of the optical slow switch does not increase significantly with the dimension. As a consequence, the power consumption of the HOS core node increases slower than the power consumption of the electronic core switch. This leads to a higher scalability of the HOS network with respect to the optical ptp network. Figure 11 also shows that when including the energy consumption of the ToR switches the gain offered by the HOS architecture is strongly reduced, highlighting again the need for a more efficient ToR switch design.

6. Conclusions

To address the limits of current ptp interconnects for data centers, in this paper, we proposed a novel optical switched interconnect based on hybrid optical switching (HOS). HOS integrates optical circuit, burst, and packet switching within the same network, so that different data center applications are mapped to the optical transport mechanism that best suits their traffic characteristics. This ensures high flexibility and efficient resource utilization. The performance of the HOS interconnect, in terms of average data loss rates and average delays, has been evaluated using event-driven network simulations. The obtained results prove that the HOS network achieves relatively low loss rates and low delays, which are suitable for today's data center applications. In particular, we suggest the use of circuits for carrying premier traffic and packets for serving best-effort traffic. Bursts can be used to provide different QoS classes, but their characteristics should be carefully designed to avoid the risk of high network delays.

The proposed HOS architecture envisages the use of two parallel optical core switches for achieving both high transmission efficiency and reduced energy consumption. Our results show that the HOS interconnect reduces by a great extent the energy consumption and GHG emissions of data center interconnects with respect to current point-to-point solutions. Furthermore, the HOS interconnect requires limited hardware modifications to existing architectures and thus can be implemented in the short/midterm and with modest investments for the operators. An open question that we plan to address in detail in future works is how efficiently

the HOS interconnect is able to scale with the increase in the servers capacity.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] G. Astfalk, "Why optical data communications and why now?" *Applied Physics A*, vol. 95, no. 4, pp. 933–940, 2009.

[2] Cisco white paper, "Cisco Global Cloud Index: Forecast and Methodology, 2011–2016," 2012.

[3] J. G. Koomey, "Worldwide electricity used in data centers," *Environmental Research Letters*, vol. 3, no. 3, Article ID 034008, 2008.

[4] J. Koomey, "Growth in data center electricity use 2005 to 2010," *The New York Times*, vol. 49, no. 3, 24 pages, 2011.

[5] http://www.google.com/about/datacenters/efficiency/internal/.

[6] Where does power go? GreenDataProject, 2008, http://www.greendataproject.org/.

[7] S. Aleksic, G. Schmid, and N. Fehratovic, "Limitations and perspectives of optically switched interconnects for large-scale data processing and storage systems," in *Proceedings of the MRS*, vol. 1438, pp. 1–12, Cambridge University Press, 2012.

[8] A. Benner, "Optical interconnect opportunities in supercomputers and high end computing," in *Proceedings of the Optical Fiber Communication Conference and Exposition (OFC '12)*, pp. 1–60, 2012.

[9] N. Fehratovic and S. Aleksic, "Power consumption and scalability of optically switched interconnects for high-capacity network elements," in *Proceedings of the Optical Fiber Communication Conference and Exposition (OFC '10)*, pp. 1–3, Los Angeles, Calif, USA, 2010.

[10] S. Aleksic and N. Fehratovic, "Requirements and limitations of optical interconnects for high-capacity network elements," in *Proceedings of the ICTON*, pp. 1–4, Munich, Germany, 2010.

[11] G. Wang, D. G. Andersen, M. Kaminsky et al., "C-Through: part-time optics in data centers," in *Proceedings of the 7th International Conference on Autonomic Computing (SIGCOMM '10)*, pp. 327–338, September 2010.

[12] N. Farrington, G. Porter, S. Radhakrishnan et al., "Helios: a hybrid electrical/optical switch architecture for modular data centers," in *Proceedings of the 7th International Conference on Autonomic Computing (SIGCOMM '10)*, pp. 339–350, September 2010.

[13] X. Ye, Y. Yin, S. J. B. Yoo, P. Mejia, R. Proietti, and V. Akella, "DOS: a scalable optical switch for datacenters," in *Proceedings of the 6th ACM/IEEE Symposium on Architectures for Networking and Communications Systems (ANCS '10)*, pp. 1–12, October 2010.

[14] A. Singla, A. Singh, K. Ramachandran, L. Xu, and Y. Zhang, "Proteus: a topology malleable data center network," in *Proceedings of the ACM SIGCOMM*, pp. 1–6, 2010.

[15] K. Xia, Y. H. Kaob, M. Yangb, and H. J. Chao, "Petabit optical switch for data center networks," Tech. Rep., Polytechnic Institute of NYU, 2010.

[16] R. Luijten, W. E. Denzel, R. R. Grzybowski, and R. Hemenway, "Optical interconnection networks: the OSMOSIS project," in *Proceedings of the 17th Annual Meeting of the IEEE Lasers and Electro-Optics Society*, pp. 563–564, November 2004.

[17] O. Liboiron-Ladouceur, I. Cerutti, P. G. Raponi, N. Andriolli, and P. Castoldi, "Energy-efficient design of a scalable optical multiplane interconnection architecture," *IEEE Journal on Selected Topics in Quantum Electronics*, vol. 17, no. 2, pp. 377–383, 2011.

[18] A. Shacham and K. Bergman, "An experimental validation of a wavelength-striped, packet switched, optical interconnection network," *Journal of Lightwave Technology*, vol. 27, no. 7, pp. 841–850, 2009.

[19] J. Luo, S. di Lucente, J. Ramirez, H. Dorren, and N. Calabretta, "Low latency and large port count optical packet switch with highly distributed control," in *Proceedings of the Optical Fiber Communication Conference and Exposition (OFC/NFOEC '12)*, pp. 1–3, 2012.

[20] C. Kachris and I. Tomkos, "Power consumption evaluation of hybrid WDM PON networks for data centers," in *Proceedings of the 16th European Conference on Networks and Optical Communications (NOC '11)*, pp. 118–121, July 2011.

[21] S. Aleksic, M. Fiorani, and M. Casoni, "Adaptive hybrid optical switching: performance and energy efficiency," *Journal of High Speed Networks*, vol. 19, no. 1, pp. 85–98, 2013.

[22] M. Fiorani, M. Casoni, and S. Aleksic, "Hybrid optical switching for energy-efficiency and QoS differentiation in core networks," *Journal of Optical Communications and Networking*, vol. 5, no. 5, pp. 484–497, 2013.

[23] M. Fiorani, M. Casoni, and S. Aleksic, "Large data center interconnects employing hybrid optical switching," in *Proceedings of the of the 18th IEEE European Conference on Network and Optical Communications (NOC '13)*, Graz, Austria, 2013.

[24] C. Kachris and I. Tomkos, "A survey on optical interconnects for data centers," *IEEE Communications Surveys and Tutorials*, vol. 14, no. 4, pp. 1021–1036, 2012.

[25] Emulex white paper, "Connectivity Solutions for the Evolving Data Center," 2011.

[26] M. Fiorani, M. Casoni, and S. Aleksic, "Performance and power consumption analysis of a hybrid optical core node," *Journal of Optical Communications and Networking*, vol. 3, no. 6, Article ID 5765579, pp. 502–513, 2011.

[27] S. Aleksi, "Analysis of power consumption in future high-capacity network nodes," *Journal of Optical Communications and Networking*, vol. 1, no. 3, Article ID 5207103, pp. 245–258, 2009.

[28] S. Kandula, S. Sengupta, A. Greenberg, P. Patel, and R. Chaiken, "The nature of data center traffic: measurements & analysis," in *Proceedings of the 9Th ACM SIGCOMM Conference on Internet Measurement Conference (IMC '09)*, pp. 202–208.

[29] T. Benson, A. Anand, A. Akella, and M. Zhang, "Understanding data center traffic characteristics," in *Proceedings of the ACM Workshop on Research on Enterprise Networking*, pp. 65–72, 2009.

[30] T. Benson, A. Akella, and D. A. Maltz, "Network traffic characteristics of data centers in the wild," in *Proceedings of the 10th Internet Measurement Conference (IMC '10)*, pp. 267–280, November 2010.

[31] "Guidelines to Defra/DECCs GHG Conversion Factors for Company Reporting," American Economic Association (AEA), 2009.