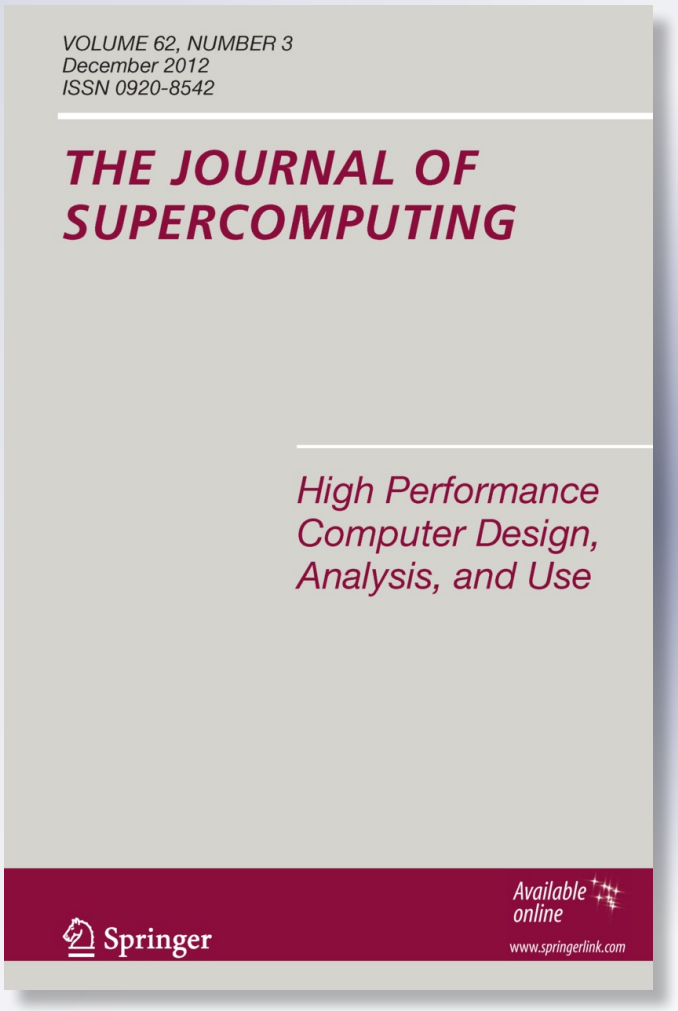
Network monitoring for energy efficiency in large-scale networks: the case of the Spanish Academic Network

José Luis García-Dorado, Eduardo Magaña, Pedro Reviriego, Mikel Izal, Daniel Morató, Juan Antonio Maestro, Javier Aracil, et al.

The Journal of Supercomputing
An International Journal of High-Performance Computer Design, Analysis, and Use

ISSN 0920-8542
Volume 62
Number 3

J Supercomput (2012) 62:1284-1304
DOI 10.1007/s11227-011-0643-z



Springer

J Supercomput (2012) 62:1284–1304
DOI 10.1007/s11227-011-0643-z

Network monitoring for energy efficiency in large-scale networks: the case of the Spanish Academic Network

José Luis García-Dorado · Eduardo Magaña · Pedro Reviriego · Mikel Izal · Daniel Morató · Juan Antonio Maestro · Javier Aracil · Jorge E. López de Vergara

Published online: 5 July 2011

Abstract The energy consumption due to information technologies is large and there are many ongoing efforts to cut it down. Previous studies have shown that there is a significant percentage of hosts which are left switched on in office buildings at night and weekend, whose energy consumption is significant. This motivates the development of techniques that can detect switched-on hosts in a simple and scalable way. This paper proposes detection techniques based on network traffic monitoring. The results show that such techniques can effectively detect switched-on hosts in small-

J.L. García-Dorado (✉) · J. Aracil · J.E. López de Vergara
Universidad Autónoma de Madrid, Francisco Tomás y Valiente, 11, 28049 Madrid, Spain
e-mail: jl.garcia@uam.es

J. Aracil
e-mail: javier.aracil@uam.es

J.E. López de Vergara
e-mail: jorge.lopez_vergara@uam.es

E. Magaña · M. Izal · D. Morató
Universidad Pública de Navarra, Pamplona, Spain

E. Magaña
e-mail: eduardo.magana@unavarra.es

M. Izal
e-mail: mikel.izal@unavarra.es

D. Morató
e-mail: daniel.morato@unavarra.es

P. Reviriego · J.A. Maestro
Universidad Antonio de Nebrija, Madrid, Spain

P. Reviriego
e-mail: previrie@nebrija.es

J.A. Maestro
e-mail: jmaestro@nebrija.es

medium campus networks and also in large country-wide networks that serve more than one million users. Interestingly, the proposed techniques can be implemented using the existing network monitoring infrastructure, specially for large networks, at a negligible additional investment.

Keywords Energy saving · Energy-aware networks · Network monitoring · Switched-on hosts · Firewall supplantation · Traffic measurement

1 Introduction

The energy consumption in the Information Technologies (IT) sector is increasing rapidly and starts to be a significant part of the energy consumption. For example, in [1], the IT related energy consumption in US offices was estimated to be around 74 TWh or 2% of the total electricity consumption in the country. In addition, this percentage is predicted to increase significantly during this decade [2]. As a result, a number of techniques have been proposed to reduce the energy consumption of computers, and networks. Most of them are based on putting elements of the system in a low power state when they are not used [3]. In [1], the savings obtained by the use of such power savings techniques were estimated to be around 23 TWh/year.

But even when power saving policies are in place, significant energy is wasted by computers that are actually not in use. For example, the authors in [1] estimated the energy savings of shutting computers down during night-time and weekends at 17 TWh/year and 7 TWh/year, respectively. This is important because many users leave their computers on at all times. However, the authors in [4] and [5] highlight the limited amount of data available for the estimation of the number of switched-on hosts during night time. In fact, hosts were checked manually in both papers, and it is suggested the need of automated mechanisms to detect activity. Such mechanisms would enable a continuous monitoring of the hosts at night and, therefore, a more precise analysis of the energy consumption and its evolution with time.

In this light, this paper provides traffic analysis techniques that enable the detection of switched-on machines in a network by monitoring hosts' network activity. Note that it is expected that almost every host connected to the Internet generates traffic because of automatic software updates (e.g., operating systems and antiviruses), background email, and VoIP clients, among other reasons, as will be shown in this paper. Once a set of switched-on machines are identified, the network managers should suggest users to switch off their computers or to use software like [6] and [7] to switch off at scheduled times. The proposed techniques can be used with either packet-level or flow-level traffic monitors. The former are common in small networks, while the latter are usually available from large wide area network routers. To the best of our knowledge, this is the first study on the use of traffic analysis for this purpose.

We have found primarily three challenges to be addressed in this approach. First, we note that firewalls are present in almost all medium-large size networks. We have found that firewalls may respond to externally originated connections on behalf of an internal host (using host's IP address), even when such host is switched off. We note that unwanted externally originated connections due to malicious traffic (e.g.,

attacks and port scans) are very common in the current Internet [8, 9]. In addition, we have found that local software firewalls running on end-hosts may block active probes. Second, flow-level measurements often suffer from packet sampling, which reduces their precision. Finally, it is possible that a host generates no traffic although it is switched on.

Our findings show that packet-level techniques can detect up to 91% of the switched-on hosts, while the Netflow-based detection techniques detect up to a reasonable 70%, even in the presence of sampling.

The rest of the paper is organized as follows. Section 2 reviews the related work. Then Sect. 3 describes the mechanisms proposed to detect switched-on hosts as well as some assumptions and limitations of our study. The next two sections are devoted to the analysis and comparison of the packet-level and flow-level approaches, at local and wide area network scales, respectively. The last section concludes this paper and provides a summary of the main findings.

2 Related work

The first studies on the energy consumption of Personal Computers (PCs) were done almost 20 years ago [10]. Nevertheless, there is a renewed interest in reducing the energy consumption of PCs. This interest was triggered by the significant amount of funds that service providers and other corporate networks spend annually on energy.

Consequently, the research community has proposed techniques to reduce the demands for energy in PCs, networked devices, and protocols. In [11], a detailed specification for PCs power management is available showing significant energy savings. In addition, low power modes have been introduced in the ADSL [12] and Ethernet [13] while other studies address energy consumption at the network level [14–16]. Note that one important point by defining such techniques is that any action has to be done in such a way that minimizes the impact on performance. To that end, prediction techniques have been proposed [17].

While all these approaches exert a significant saving in terms of energy and money, the authors in [1] showed that an extensive set of hosts, which are actually not in use, remain on at weekends and nighttimes, and consequently wasting energy. Studies conducted in academic and research institutions [4] showed a remarkable power consumption due to this inefficient use of computers. There are ample variations in the percentage of hosts that are left switched-on at night ranging from 35% in [5] to 63% in [4], in both cases verified with manual inspection. Therefore, we aim at filling this gap by showing how to automatically identify such a set of hosts analyzing network activity.

The authors in [6] presented an open-source software that allows users to schedule power states of their PCs. The results showed that this tool reduces the average PC uptime about 40%. Similarly, commercial products [7] are becoming available to monitor and manage the energy consumption of networked PCs after installing proprietary software. Both tools are complementary to the techniques proposed in this paper. Essentially, once a host is detected switched-on during some nights or weekends, the network managers should suggest users to install software such as the above mentioned to save energy.

Regarding the identification of the switched-on host, the authors in [6] leverage "port scan" mechanisms to estimate the number of switched-on PCs in a university campus over 24 hours a day. However, such approach presents some limitations from both accuracy and scalability viewpoints. First, we have found that the use of "port scan" mechanisms provide clearly worse results than other active probing techniques because of local software firewalls running on end-hosts. Additionally, such approach injects proactively traffic to any host in the network, which is not a scalable solution for wide area networks.

3 General solution and analysis scenarios

This paper proposes to automatically detect switched-on hosts analyzing activity through network monitoring using both packet-level and flow-level traffic monitors. To evaluate the accuracy of the proposed techniques, we use measurements from two different scenarios: the Spanish Academic Network (RedIRIS) [18], and a selected campus within RedIRIS, the Public University of Navarra (UPNA) campus. The traffic collection in RedIRIS is at the flow record level (Netflow) whereas packet level analysis is performed in the UPNA campus network. Figure 1 shows the RedIRIS topology and the UPNA location inside the RedIRIS network, as well as their measurement system architectures.

RedIRIS provides advanced communication services to the scientific community, hospitals, and to most Spanish campuses, which comprises over 350 affiliated institutions and 18 Points of Presence (PoP) across the country. For such a large network, we pursue detection techniques that are cost-efficient and pay off for the sav-

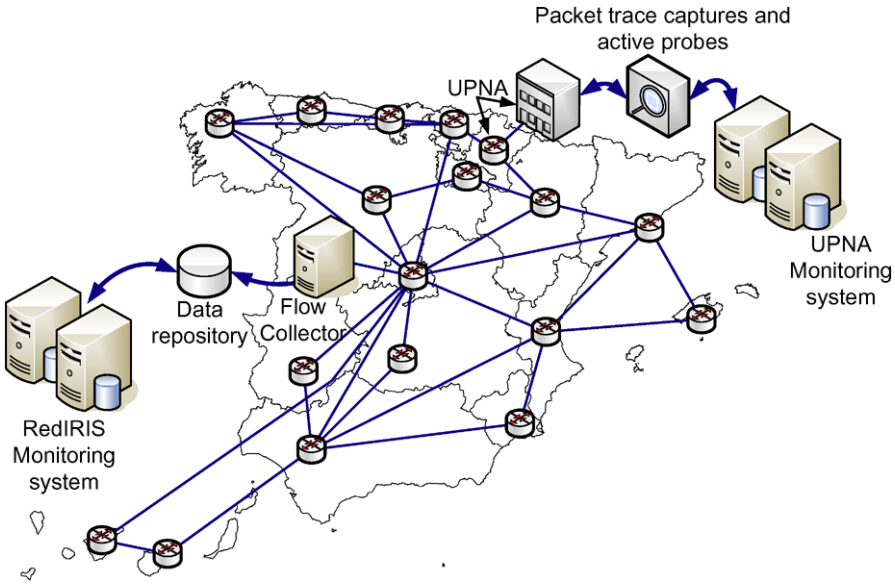


Fig. 1 RedIRIS network topology and measurement system architectures of RedIRIS and UPNA

ings in energy consumption. To do so, we detect activity through analyzing Netflow records [19], which are available in the majority of access routers, with no need of installing additional network probes, making the hardware investment negligible. Therefore, the only Capital Expenditure (CAPEX) is that of a server that collects the Netflow records. Netflow records provide a summary per flow, either UDP or TCP, which includes the source and destination IP addresses, ports, and number of bytes and packets among others. For a trace duration of 3 years, inexpensive hard disks and processing servers have been used. Needless to say, there is no need to collect 3 years worth of data to detect activity, which can be done on a day-by-day basis, with a very low hardware investment. In addition, the bandwidth consumption due to the collection of Netflow records is not significant. For the whole Spanish academic network, the Netflow records worth of throughput is 2 Mb/s in mean, which is very little for a Gb/s backbone network. Furthermore, the Netflow records can also be used for generic traffic monitoring purposes, beyond detection of switched-on hosts. Finally, it is also worth noting that RedIRIS routers carry out packet sampling for Netflow, with a sampling rate of one out of 100 packets [20].

On the other hand, UPNA is a mid-size university with some 8,000 users, including faculty, staff, and students. The access links to RedIRIS have 1 Gb/s capacity. We installed electrical splitters in the access links that redirect a copy of each packet to the measurement system. Clearly, packet-level detection techniques are more accurate than their Netflow-based counterpart. However, this is at the expense of a more expensive measurement infrastructure and larger computational cost. Note that this is not an option for the whole Spanish academic network because the amount of data collected would be humongous. To further assess the accuracy of the packet and flow based detection techniques, we proactively sent ARP queries to all the UPNA hosts and we compared the ARP results with those from packet-level and flow-level analysis. Then we applied the lessons learned from the study of UPNA measurements to a more extensive set of RedIRIS' institutions and finally to the whole RedIRIS network.

The proposed detection techniques can be applied to a wide range of power management policies, as defined by the network manager. For example, a simple power management script may send an email to those users whose machines have been detected as switched on during consecutive days, from 00:00 a.m. to 6:00 a.m., suggesting switching off their machines or using specific software to switch them off at scheduled times (for instance, [6] and [7]). It will be shown that very significant power savings can be achieved, which compensate for the extra effort of implementing such policies.

3.1 Assumptions and limitations

We assume that a host that is sending traffic after office hours is not active (it is not purposely running some process) which may not be true. For example, a computer simulation may be running overnight and we will tell that the computer is switched on. It is the network manager duty to judge whether a detected switched-on host is in good standing. For instance, mail and web servers will be also detected switched on, but the network manager can filter them out easily because their IP addresses are typically known.

On the other hand, the proposed detection techniques are not compatible with Network Address Translation (NAT) mechanisms. Should NAT be in use, our detection techniques would provide a lower bound of estimated energy consumption, because some hosts would be "hidden" behind a single IP address. Detection techniques to estimate the number of hosts behind a NAT [21] could be eventually used in specific networks. In those techniques, the IP identifier (IPid) field in IP header is used to identify packets belonging to different hosts even with the same source IP address. A host uses an incremental counter to fill up the IPid for each packet in order to differentiate packets in reception in case of fragmentation. The evolution of this counter is different for each host depending on the rate of IP packets generated and the initial value of IPid. Therefore, this behavior can be used to differentiate hosts hidden by the same NAT-IP address. However, note that such methodology not only requires to capture any packet that traverses a given network but also the analysis of their headers. Such extra load is not scalable for wide area networks.

In addition, note that by measuring in the access links of the RedIRIS' institutions, a switched-on host will be detected if and only if it generates traffic to the Internet, not only internal traffic. A few simple experiments confirm that the majority of hosts generate traffic to the Internet. For example, it was observed that most Linux and Windows machines periodically check for updates (both due to operating system and software applications like antivirus, Java, Office, among others). Besides, many users tend to leave some applications running, such as Skype and email clients. Finally, also some application protocols, such as Network Time Protocol, generate traffic periodically.

As a conclusion, it should be possible to detect a significant number of hosts even if measurements are only performed in the access link. The results presented in this the paper confirm such hypothesis.

Finally, we faced two important challenges. First, hosts generating a small number of packets may be left undetected because of the Netflow packet sampling (in this case, one out of 100 packets are sampled). Second, firewalls may respond to externally originated connections on behalf of the internal host, thus increasing the false positive ratio. The proposed detection techniques have been designed to cope with these issues.

4 UPNA campus network

In this section, the detection of switched-on hosts in the UPNA campus area network is addressed. To this end, we performed active probing inside the campus and passive monitoring at packet and flow level in its access link to RedIRIS. For the sake of clarity, the following results correspond to the night (00:00–06:00 a.m.) of May 13, 2009, although we remark that equivalent results were obtained over the months of May and June 2009.

4.1 Active probing techniques

Clearly, the best we can do to identify switched-on hosts is to send probe packets to each of the IP addresses in the network range. An adequate choice of probe packets produces replies from the targeted hosts.

Among the available probe packets, the ICMP Echo-based [22] are the most popular (Ping tool). A central probing station sends ICMP Echo Request probes that should be replied with ICMP Echo Reply packets from each switched-on host. This reply is generated by the operating system. However, local software firewalls may block such replies. Typically, software firewalls block incoming traffic not produced by the user, and sometimes the ICMP Echo Request probes are also dropped.

As an alternative, we may use specific applications that scan services running in a certain IP range. Nmap [23] is an example of a network security and auditing tool that can check if a given host is switched on by trying connections to typical TCP/UDP ports. However, the experimental results are not good enough, again because of local software firewalls running on end-hosts.

Both ICMP Echo-based and Nmap-based approaches can be launched from any probing station in the Internet with IP connectivity with the target network. However, in our case, we can also probe hosts at link level, the best option being the Address Resolution Protocol (ARP) request/response packets. Nowadays, most LANs are Ethernet based and hosts are enforced to respond to ARP request packets. In addition, it is worth remarking that switched-off hosts equipped with Wake-on-LAN capability do not respond to such requests and that there is not any switch working as ARP-proxy in UPNA. Thus, a probing station inside the LAN can use the ARPing tool [24] which provides this probing functionality and it is not affected by the presence of local software firewalls in the end-host. The experimental results provide a total number of 528 hosts responding to ARPing during the night of May 13, 2009. During office hours, 1,222 switched-on hosts were detected using ARPing tool. This means that 43% of hosts were left switched on during that night. Such 528 hosts are switched on for sure. The ICMP Echo-based approach identifies 476 switched-on hosts (90%) and the Nmap-based approach only identifies 302 switched-on hosts (57%).

However, the ARP-based approach has severe drawbacks because the probing station must be inside the target LAN and it is proactively injecting traffic. This is not a scalable solution for a wide area network such as RedIRIS. Thus, the results provided by the ARP-based approach will only be used as a reference to evaluate the following approaches which are aimed to provide a more scalable and lower cost solution. Table 1 shows the results for the set of proposed methodologies to identify switched-on hosts which will be explained in the following sections. Additionally, Fig. 2 shows the results as a Receiver Operating Characteristic (ROC) space to facilitate the visual comparison.

In the subsequent analysis, the following metrics and their respective rates will be considered to compare diverse approaches:

- true positives (TP): hosts that are detected switched on that agree with the ARP-based approach.
- true negatives (TN): hosts that are detected switched off that agree with the ARP-based approach.
- false positives (FP): hosts that are detected switched on that disagree with the ARP-based approach.
- false negatives (FN): hosts that are detected switched off that disagree with the ARP-based approach.

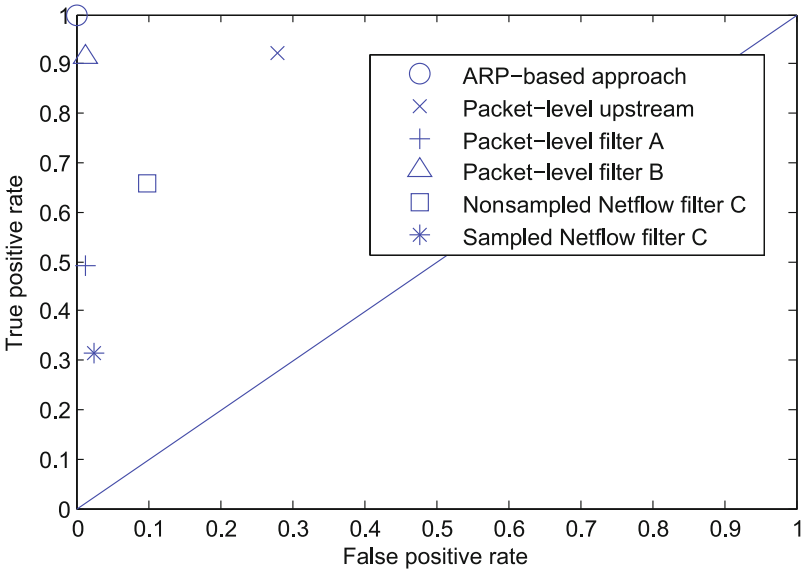


Fig. 2 The ROC space and plots of the different approaches for UPNA campus network

Table 1 Results for UPNA campus network

Technique	TP	TN	FP	FN
ARP-based approach (ref.)	528	694	–	–
Packet-level upstream	486	500	194	42
Packet-level filter A and Nonsampled Netflow filter A	259	685	9	269
Packet-level filter B	483	685	9	45
Nonsampled Netflow filter C	348	626	68	180
Sampled Netflow filter C	166	676	18	362

4.2 On the effect of firewalls

Passive monitoring at packet level in the Internet access link provides valuable information to detect host activity. The most straightforward approach considers that a host that generates upstream traffic is switched on. In fact, at a first glance, a given IP address can only generate packets if and only if the corresponding host is switched on. However, the results presented in Table 1 as "Packet-level upstream" serve to reject that assumption. In that row, all IP addresses that generate upstream packets are considered as active. Surprisingly, the number of false positives is very significant.

The false positives are primarily due to a firewall, that is breaking the end-to-end Internet paradigm. This is a stateful and transparent firewall that keeps track of the state of network connections going across [25]. In fact, incoming TCP connections from Internet are received and answered by this firewall. The firewall completes the

TCP three-way handshake, accepting connections for target hosts that are not even reachable (in our case switched off). Packets generated by the firewall keep the original internal IP address. Thus, the firewall-generated traffic cannot be distinguished from the internal host generated traffic. This behavior is useful in order to protect the network from external attacks (Denial of Service attacks, for example). Only the nonsuspicious connections are passed through the firewall to the internal host. Firewalls are present in almost all medium-large size networks and the stateful behavior is typical, this clearly affects detection techniques based on traffic monitoring. In our case study, the firewall acts over incoming TCP connections directed to well-known ports (web, ssh, smtp, etc.) and not over the rest of connections. Therefore, the presence of upstream traffic from a certain IP address does not necessarily imply that the host with that IP address is active. Many of those incoming TCP connections are network attacks that try to scan several IP addresses within the subnetwork. In this regard, the authors in [8] showed that the 16-bit-address ranges of two universities were scanned in their entirety ($2\times65{,}534$) in only a 27-hour trace (20 minutes traces, four times a day, 20 days). In conclusion, the number of false positives (194 hosts, 37%) is not acceptable since each false positive may result in a notification to a user that actually does not left any host switched on during nighttime. In this light, the following approaches aim at reducing the false positive rate to avoid disturbing the users unnecessarily, albeit often this is at the expense of increasing the number of false negatives.

To remove the firewall effect, two different approaches are proposed. The first one (labeled I in Fig. 3) only considers flows initiated by one of the UPNA IP addresses. It ignores the rest of connections (hosts that reply to externally initiated flows) due to the possible firewall supplantation. Let us refer to this methodology as "campus sourced traffic detection methodology." On the other hand, the second approach (labeled II in Fig. 3) tries to identify the firewall connections and rules them out. The UPNA's firewall replies can be detected because they follow the TCP three-way handshake and the source port is well known. Furthermore, the connection payload contains 0 bytes of TCP data. Several other alternatives have also been considered, which take into account the number of packets interchanged with certain TCP flags, with any possible source TCP port. However, they provide worse performance in terms of TP

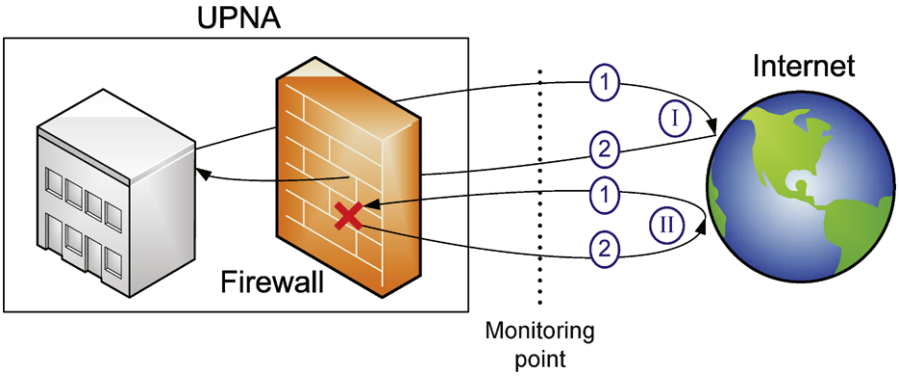


Fig. 3 Techniques for removing the firewall effect

and FP rates. The downside of this technique is that it requires knowledge of the firewall behavior, which may be different depending on the model and configuration. In what follows, this methodology is denoted by "firewall detection methodology."

4.3 Packet-level techniques

First, the campus sourced traffic detection methodology will be applied to packet-level measurements during the night under analysis. This approach is called "Packet-level filter A" and it requires to keep track of packets belonging to flows with common source/destination IP addresses and ports. In this case, we consider the switched-on hosts are those with at least one UDP flow or TCP connection originating from the internal network to the Internet. Particularly, in the case of TCP flows, it is straightforward to control packets with SYN flag activated and determinate if the flow was indeed begun by a local hosts. On the other hand, the UDP version of the method requires to keep track of the active UDP flows in order to identify if a given packet is the first one of its flow and the institution's local IP ranges to determine its direction. Note that this methodology can also be performed with flow-level monitoring information at a lower cost as discussed in the next section.

Figure 4 shows, as pseudocode, the implementation of this technique. The result consists of a report that contains as a list the candidate hosts to be switched off at nighttime once removing the IP addresses of the known servers. As stated before, the network manager could, for example, contact users of such hosts by email suggesting turning off their machines or using specific software that automatically switches off the computer at scheduled times.

The results presented in Table 1 for "Packet-level filter A" show some 48% of true positives. This indicates that there are a large number of switched-on hosts that only generate traffic in response to externally initiated connections (for instance, ssh logons or database queries) that are not discovered by this technique. It is worth noting that some of such externally initiated connections correspond to scanning attacks.

```
Initializate switched_on_hosts_list
Initializate active_UDP_flows_list
Initializate local_IP_list
Initializate servers_list
for each packet do
  if protocol is TCP and only flag SYN is activated and IP_source is in local_IP_list and
  IP_source is not in switched_on_hosts_list then
    add IP_source in switched_on_hosts_list
else if protocol is UDP and IP_source is not in active_UDP_flows_list and IP_source is in
      local_IP_ list and IP_source is not in switched_on_hosts_list then
      add IP_source in switched_on_hosts_list
      update active_UDP_flows_list
  end if
end for
report_list ← switched_on_hosts_list not in servers_list
return report_list
```

Fig. 4 Packet-level filter A technique implementation

```
Initializate switched_on_hosts_list
Initializate servers_list
for each upstream packet do
  if protocol is TCP and source_port < 1024 and TCP_ data = 0 then
    continue   else if IP_source is not in switched_on_hosts_list then
    add IP_source in switched_on_hosts_list
  end if
end for
report_list ← switched_on_hosts_list not in servers_list
return report_list
```

Fig. 5 Packet-level filter B technique implementation

This finding constrains the applicability of the detection technique. Conversely, this technique is advantageous because it can be applied to any network regardless of the possible firewall operation. Additionally, it shows a low number of false positives (1.5%).

To increase the detection rate, the firewall detection methodology is applied. In this case, we consider all hosts that generate traffic as switched on, either if it is internally or externally initiated. We consider firewall supplantation and filter out those hosts that generate only TCP traffic, from well-known ports, with 0 bytes total TCP data payload. The results are provided in Table 1, "Packet-level filter B," showing that 91.5% of switched-on hosts are detected correctly. The number of false positives (1.5%) and false negatives (8.5%) shows a good trade-off between both ratios. Moreover, the implementation of this technique is as simple as Fig. 5 shows.

Note that the processing at packet-level could not scale for large networks. Consequently, flow-level detection techniques are considered in the next section.

4.4 Flow-level techniques

In this section, the results of applying both proposed methodologies to Netflow records are shown. The firewall detection methodology can be applied to sampled and nonsampled Netflow. On the contrary, the campus sourced traffic detection methodology can only be applied to nonsampled Netflow records. Note that it is not possible to determine which side initiated a connection in presence of sampling. More specifically, if the first packet is not sampled, there are no means to identify which host initiated the connection. In addition, note that the identification of hosts which initiated a connection can be carried out using either nonsampled Netflow records or packet traces with no difference in the results, as shown in Table 1 ("Nonsampled Netflow filter A"), but cutting the computational cost as the analysis is done per flow and not per packet as Fig. 6 shows.

Regarding the "firewall detection methodology," we apply the filter B introduced in Sect. 4.3 with some modification to adapt it to the case of the Netflow records, both nonsampled and sampled Netflow records:

In the nonsampled Netflow case, Netflow aggregates the total number of bytes (including headers) of consecutive packets that share the same 5-tuple (IP addresses, ports, and protocol) for each direction. Thus, it is not possible to identify if there

```
Initialize switched_on_hosts_list
Initialize local_IP_list
Initialize servers_list
for each Netflow do
  if IP_source is in local_IP_list and IP_source is not in switched_on_hosts_list then
    add IP_source in switched_on_hosts_list
  end if
end for
report_list ← switched_on_hosts_list not in servers_list
return report_list
```

Fig. 6 Nonsampled Netflow filter A technique implementation

```
Initialize switched_on_hosts_list
Initialize local_IP_list
Initialize servers_list
for each flow do
  if protocol is TCP and source_port < 1024 and average_packet_size_bytes < 65 then
    continue
  else
    if IP_source is in local_IP_list then
      IP_local ← IP_source
    else
      IP_local ← IP_destination
    end if
    if IP_local is not in switched_on_hosts_list then
      add IP_local in switched_on_hosts_list
    end if
  end if
end for
report_list ← switched_on_hosts_list not in servers_list
return report_list
```

Fig. 7 Nonsampled and sampled Netflow filter C techniques implementation

are 0 bytes of TCP data as requested by filter B. We adapt the filter to one that uses an estimation of the mean packet size to identify TCP flows without TCP data. It is difficult to make an accurate estimation of the packet size which is required by the filter. Nevertheless, the average packet size, namely the ratio between the number of bytes and packets, serves as an estimate. Considering that packet sizes also include the IP header, the minimum TCP packet size is 40 bytes (20 bytes basic IP header + 20 bytes basic TCP header). However, optional fields can be used at IP and TCP level, increasing the minimum packet size. Such approximation may not be exact as the firewall response includes packets with different sizes due to the TCP/IP headers, which may also vary in size. Therefore, we consider that a firewall response is a flow with average packet size smaller than 65 bytes. Such threshold value has proven to be the best trade-off between the true and false positive rates. The results are shown as "Nonsampled Netflow filter C" in Table 1 and Fig. 7 details its implementation. The number of hosts identified correctly is 348, which is less than the packet-level

technique. On the other hand, there are a number of false positives that indicate that some firewall responses are being classified as traffic from a switched-on host.

In the other case, sampled Netflow, Table 1 shows the results of using UPNA's sampled Netflow records at a 1:100 rate in label "Sampled Netflow filter C." These results are useful to evaluate the impact of the sampling process in filter C. We note that this is the typical sampling ratio of the RedIRIS' routers. Filter C is robust against packet sampling, because all packets in a firewall response do not have payload, and consequently a firewall flow has a small average size value, regardless of possible sampling. However, some nonfirewall responses may be due to the random sampling process that captures sizes lower than 65 bytes, even though the connection includes larger packets. By comparing these results to previous sections, it becomes apparent that packet sampling has a significant impact. It is worth noticing that the number of IP addresses detected (i.e., the number of IP addresses for which at least one packet was sampled) was 275, this is roughly equal to half the actual number of switched-on hosts (528). However, false positives, which represent the number of unnecessary alerts, still remains within reasonable values. This implies that filter C is robust against the firewall even though the sampling effect prevents us from detecting the whole set of switched-on hosts.

Actually, some switched-on hosts generate such a small amount of packets that none of them are sampled by the Netflow agent at the router. In order to mitigate such effect, a longer measurement period is considered, therefore, the process to identify switched-on hosts should not be launched every day but at a longer period of time. Specifically, the entire month of June 2009 during the night is analyzed. We support the hypothesis that an important fraction of the hosts are left switched on night after night, Fig. 8 confirms this point. It shows as a histogram the number of days that each of the hosts left switched on the first day of the measurement campaign (1 June 2009) remained switched on during the entire month under study according to ARP-based approach. More than 50% hosts were switched on during the 30 days of June, above 20% hosts were on between 2 and 4 weeks, and only 10% were turned off in the first five days. In this light, we increase the chances of detection of the hosts that usually remain switched on at nighttime by extending the measurement period.

Figure 9 shows the results of applying the firewall detection methodology in sampled Netflow data to the entire month of June 2009. Bearing in mind the results of Fig. 8, which showed that typically switched-on hosts at nighttime remain on during the most of the 30 days under study, we have considered that one host is switched on if at least one night it was detected as so. As shown, the ratio of detected hosts increases monotonically as more days are added to the sample. Furthermore, the number of false positives remains negligible during all the experiment. At the aggregation level of one week, more than 50% of the total switched-on hosts were detected and about 4% of the detections were false. More specifically, 360 hosts were correctly detected out of 674 hosts detected by ARP ping; the number of false detections is only 10. By analyzing the entire month, we found that 70% hosts can be detected with similar ratio of false detection. In this case, 660 hosts were correctly detected (out of 943) and 22 were incorrectly identified as switched on. All this information could be stored in a DDBB to facilitate its management, but note that it must be updated to detect changes on the behavior of the users. That is, users that start to switch

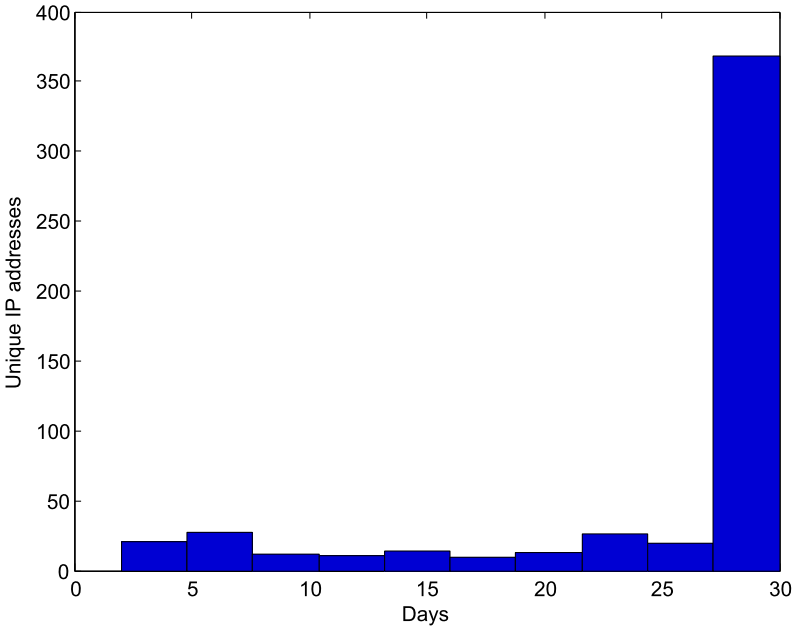


Fig. 8 Number of days that the hosts switched on by June 1, 2009, remained on in the entire month of June 2009

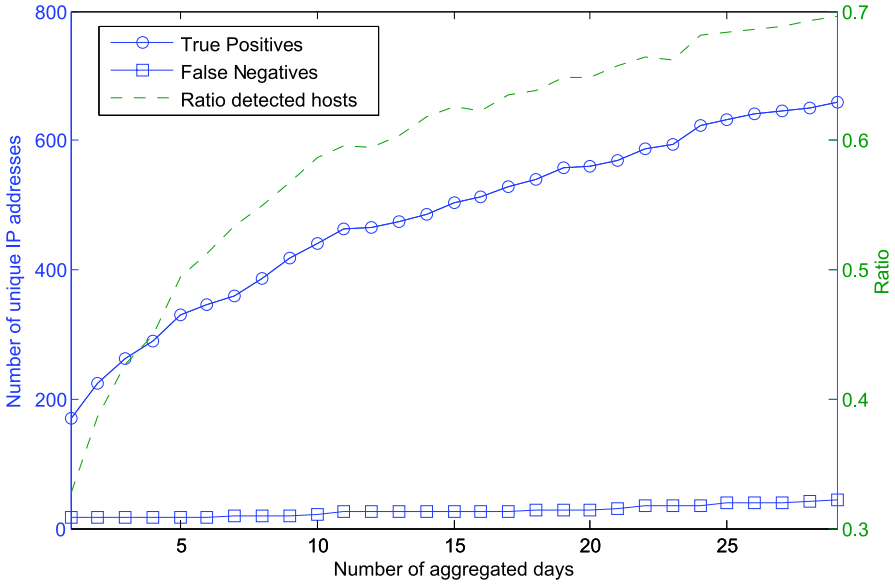


Fig. 9 True positives and false positives along with the ratio of detected hosts for one month worth of data

off their computers. We believe that an interval in the range of one or some weeks is an appropriate period of time to warn users that their computer remains on during nighttime.

5 Case study: RedIRIS

The previous sections have presented different approaches for the detection of switched-on hosts for a campus network. Such approaches include ARP probing, packet-level, and flow-level analysis, with and without packet sampling. For large network backbones such as RedIRIS, neither ARP probing nor packet-level analysis are a choice. First, ARP probing requires link level accessibility, however, RedIRIS, as happens in most large networks, does not have access to any of the internal routers of its institutions. Second, the required computational and storage capacity makes packet-level analysis unfeasible. Similarly, the router load increases if nonsampled Netflow is in use, and routing may be compromised. Consequently, the most cost-effective technique for RedIRIS is sampled Netflow. Actually, RedIRIS' routers are configured with a Netflow sampling rate of 1:100, for many years. This is less demanding in terms of computational load and infrastructure requirements but it is also less accurate.

In this section, we extend our findings to the whole academic network, with the final aim to estimate the total number of switched-on hosts in RedIRIS. So far, we have focused on a single university (UPNA) network and have found that the application of the firewall detection methodology to sampled Netflow provides a lower bound to the number of switched-on hosts. As will be shown, such lower bound represents a remarkable number of hosts which motivates the application of this technique to the whole country-wide academic network. First, we extend the results to a set of similar campus networks, in order to assess the accuracy. Then the analysis for the whole RedIRIS is presented.

5.1 Lower bound to the number of switched-on hosts in an extensive set of campus networks

In order to compare a homogeneous set of networks, 9 universities have been carefully selected out of the total set of RedIRIS institutions, which share NAT, Proxies, and P2P policies. Such homogeneity is important because, for instance, the use of NAT could artificially reduce the number of IP addresses that access the Internet and the estimated number of switched-on hosts. Similarly, limitations to the use of P2P applications can discourage users from using their computers. All the selected centers make negligible use of NAT and proxies, if any, and they do not impose restrictions to P2P applications. The networks' population sizes ranges from 8,000 up to 50,000 users.

We focus on the space and time diversity [26] of the number of switched-on hosts. More specifically, we analyze whether the number of switched-on host candidates overnight is large enough to motivate the implementation of any energy saving technique. Accordingly, we show the detected switched-on hosts in absolute and relative

figures over the total number of switched-on hosts during daytime. In addition, we show the results for a week worth of data to assess if the estimations remain stable over time.

We have applied the firewall detection methodology (Sect. 4.4) to the set of 9 campus networks during June 2009. We note that RedIRIS provided us with sampled Netflow records and the results of the firewall detection methodology were only validated in the UPNA campus network. However, it does not render the analysis useless as: (i) the number of IP addresses detected in UPNA using sampled Netflow records was lower than the actual number of switched-on hosts, during all the measurement campaign; this suggests that the impact of sampling is more important than the firewall effect. The detection percentage for switched-on hosts was 52% with sampling and 91% with filter of firewall effect (filter B). (ii) In case other firewalls do not respond to certain external connections, we would filter more IP addresses than we should. Consequently, it becomes evident that the firewall detection methodology gives a lower bound to the actual number of switched-on hosts.

In this light, Fig. 10 shows the estimated number of switched-on hosts during the first week of June 2009 at nighttime (from 0:00 a.m. to 6:00 a.m.) and also during the same period at daytime (from 12:00 p.m. to 6:00 p.m.) for the set of campus networks (labeled as $U_1 \ldots U_9$). The dashed-line represents the ratio between such two values. As shown, such ratio ranges from 10% to 37% from Monday through Friday in all networks under study and it is even larger during the weekend. In absolute terms, it involves about 10,000 switched-on hosts in the set of institutions under analysis per day. This result shows that the number of hosts that are switched on at night is significant and reinforces the interest of this work.

Additionally, this last figure also shows how the human-activity dynamics [27] arises. Essentially, weekdays show similar behavior between them with a stable night/day ratio. However, during the weekend, the number of switched-on hosts in the mornings dips to such an extent that overlaps with the number of switched-on hosts at nights. Such behavior was found during all the measurement campaign, but, for the sake of brevity, only a week worth of data is presented.

5.2 Switched-on hosts at nights in whole RedIRIS network

The previous sections have shown the analysis of a specific campus network (UPNA) and its extension to a set of similar networks. This section takes one step further and extends the study to the whole RedIRIS network. Recall that RedIRIS includes more than 350 institutions with more than 1,000,000 users. In this case, the analysis of the switched-on hosts during the nighttime has a special interest since the potential energy savings are even more significant.

Figure 11 shows the number of switched-on hosts in RedIRIS network for the first week of June 2009 according to the firewall detection methodology with sampled Netflow records. As can be seen, the number of candidate hosts to be switched off is larger than 40,000. This includes more than a quarter of the total number of IP addresses with activity in RedIRIS network, roughly 150,000 hosts. It is worth noting that a small fraction of the RedIRIS' institutions use NAT. For such institutions, the use of NAT results in a reduction of the estimated number of switched-on hosts.

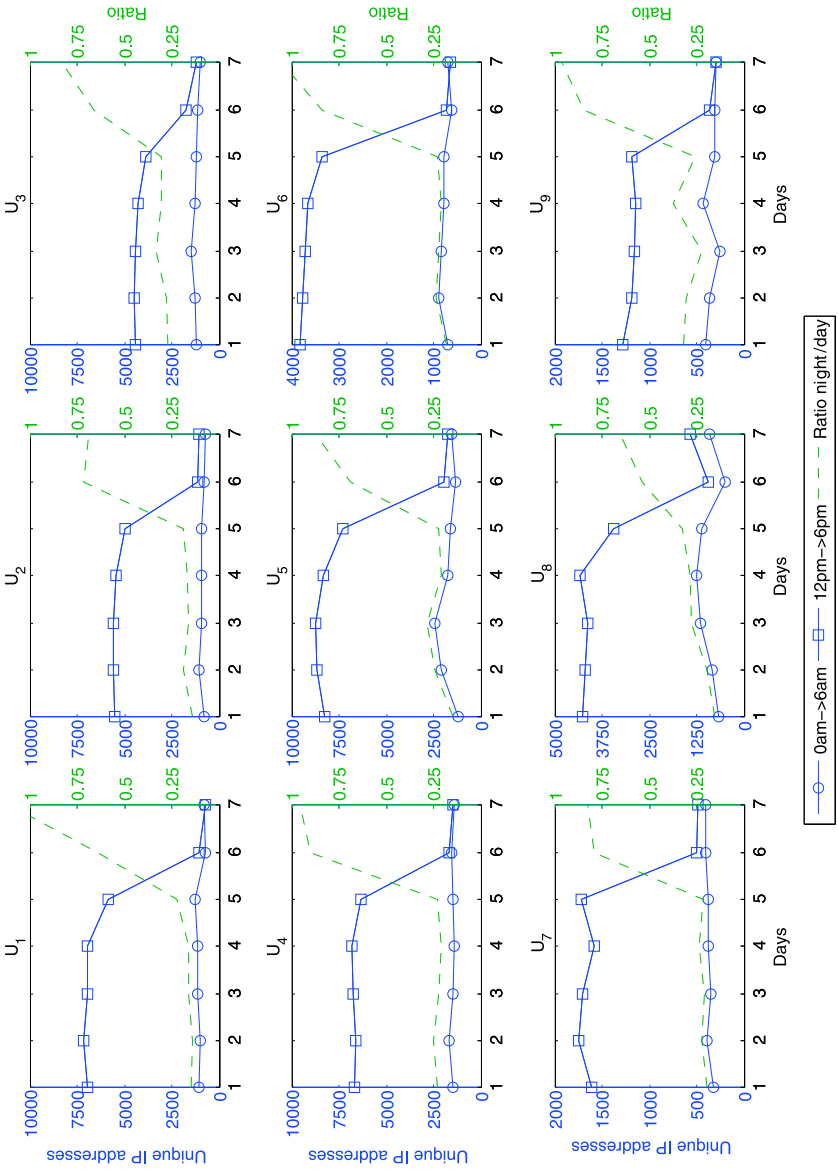


Fig. 10 Number of unique IP addresses marked as switched-on hosts at daytime, nighttime, and ratio for a set of 9 campus networks for a week

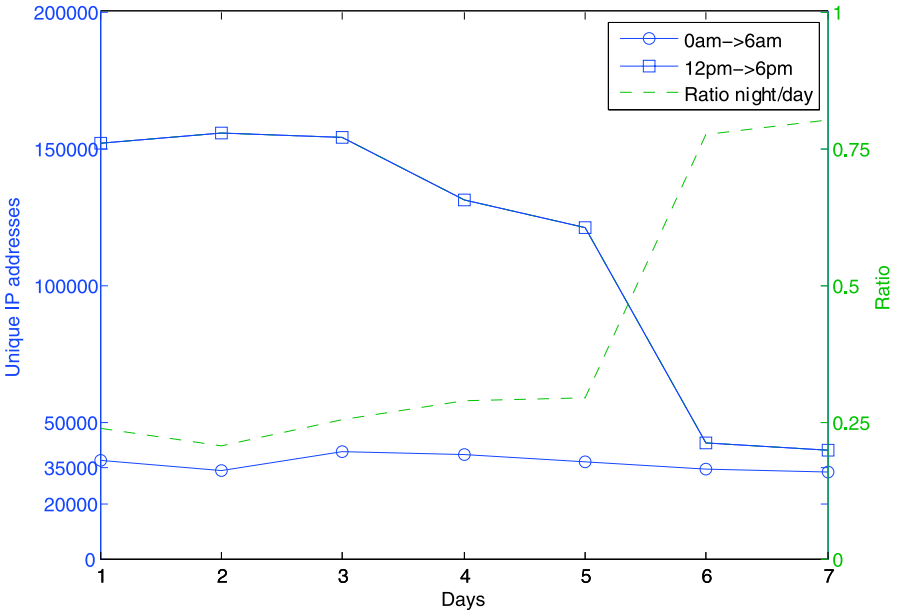


Fig. 11 Number of unique IP addresses marked as switched-on hosts in RedIRIS

Therefore, it makes the estimation of the number of switched-on hosts even more conservative.

The power consumption of a host can vary substantially from one host to another and for the same host depending on the power saving options in use [5]. As a conservative estimate, assuming that most hosts have advanced power saving options enabled and use a 10 W power consumption per host, we would obtain 400 kW for the 40,000 hosts. If the hosts are unnecessarily on for 12 hours a day, that would mean a consumption of 4,800 kWh per day. This would correspond to 1,752 MWh per year with an estimated cost of over 250 k Euros per year (assuming 0.15 Euros per kWh) and emissions of over 1,000 tons of CO_2 (using the average CO_2 emissions per kWh in [28]).

As mentioned before, this is a conservative estimate; let us consider the typical nighttime energy consumption of 200 kWh reported in [5] per host per year when the host is on at nights and spends most time in low power mode. The number of hosts left on at nights is approximately equal to 40,000, therefore, we obtain some 8,000 MWh as a result, with a cost of 1.2 million Euros per year and CO_2 emissions of more than 4,666 tons.

To put these figures in perspective, the number of households that would consume the same amount of energy is estimated. Assuming an average consumption of around 4,000 kWh per household [29], the energy wasted would be equivalent to more than 400 households for the conservative estimate. This is a substantial amount of energy and money that is currently wasted by universities and could be easily saved.

6 Conclusions

In this paper, we have proposed and analyzed several switched-on host detection techniques for energy efficiency in large scale networks based on network monitoring. Such techniques have been presented in a top-down approach from those that require some local accessibility (ARP-based) or relative high computational burden (packet inspection) to less-resource demanding techniques as Netflow. More specifically, the ARP-based solution would need a computer per network segment, or if possible, a computer connected to each network segment. Packet-level solutions would require a computer per access network. This computer should have enough computational power to analyze all the packets traversing the network. Nonsampled Netflow solutions would also require a computer per network, but with less computational power, given that dealing with flows reduces the number of analysis. Finally, the less costly solution would be the sampled Netflow, which is provided by most already deployed routers. However, this reduced investment and complexity is at the expense of precision (Fig. 2). Nonsampled Netflow technique based on traffic direction is a more general approach but its performance can be clearly improved by techniques based on the detection of the firewall responses. Finally, sampled Netflow have turned out to be a good compromise between cost (often negligible) and accuracy. Moreover, extending the measurement campaign duration improves the precision notably. In this light, Fig. 12 shows graphically all these techniques according to their cost and general accuracy.

In addition, we have shown that the number of switched-on hosts during nights is a shared characteristic of an extensive set of networks during a representative period of time. Given the large number of switched-on hosts detected and the evident benefits of saving energy, the proposed techniques could be useful not only in academic networks, but also in corporate networks that are geographically disperse such as for example, banking networks. However, the limitations discussed in Sect. 3.1 should be taken into account to ensure applicability.

In conclusion, we have provided a number of detection techniques for switched-on hosts that apply to a wide variety of network setups, according to the resources and

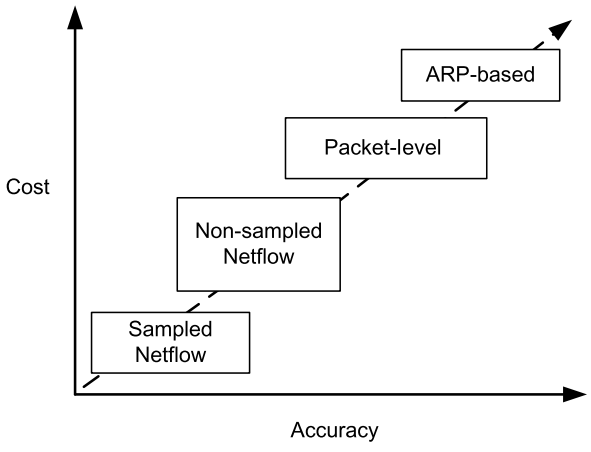


Fig. 12 Relationship between the cost of methodologies and the ratio of switched-on hosts we could positively identify

monitoring equipment available. Specifically, in the case of Netflow, we have shown that the energy bill and pollution of a large scale network can be easily reduced with a marginal extra expenditure.

Acknowledgements This work has been partially funded by the Spanish Ministry of Education and Science under project *ANFORA* (TEC2009-13385) and the F.P.I. Fellowship program of Spain. The authors would like to acknowledge the support of the National Research and Education Network, RedIRIS. Finally, we would also like to thank the anonymous reviewers who helped us to improve the quality of the paper.

References

1. Kawamoto K, Koomey JG, Nordman B, Brown RE, Piette MA, Ting M, Meier AK (2001) Electricity used by office equipment and network equipment in the U.S. LBNL-45917 Report, Environmental Energy Technologies Division, Ernest Orlando Lawrence Berkeley National Laboratory University of California, Berkeley, USA, February
2. United States Report Addendum (2008) Smart 2020: Enabling the low carbon economy in the information age. A report by The Climate Group on behalf of the Global eSustainability Initiative
3. Benini L, Bogliolo A, Micheli GD (2000) A survey of design techniques for system level dynamic power management. IEEE Trans Very Large Scale Integr 8(3):299–316
4. Roberson JA, Webber CA, McWhinney MC, Brown RE, Pinckard MJ, Busch JF (2004) After-hours power status of office equipment and energy use of miscellaneous plug-load equipment. LBNL-53729 Report, Environmental Energy Technologies Division, Ernest Orlando Lawrence Berkeley National Laboratory University of California, Berkeley, USA, May
5. Nordman B, Meier A, Piette MA (1998) PC and monitor night status: power management enabling and manual turn-off. LBNL-46099 Report, Environmental Energy Technologies Division, Ernest Orlando Lawrence Berkeley National Laboratory University of California, Berkeley, USA, July
6. Chiaraviglio L, Mellia M (2010) Polisave: efficient power management of campus PCs. In: Proceedings IEEE international conference on software, telecommunications and computer networks, Bol, Croatia, September, pp 1–6
7. Verdiem Surveyor (2011) http://www.verdiem.com
8. John W, Tafvelin S (2007) Differences between in- and outbound Internet backbone traffic. In: Electronic proceedings of Terena networking conference, Copenhagen, Denmark, May
9. Ricciato F (2006) Unwanted traffic in 3G networks. ACM SIGCOMM Comput Commun Rev 36:53–56
10. Newsham G, Tiller D (1992) Case study of the energy consumption of desktop computers. In: Proceedings of the IEEE industry applications society annual conference, Houston, USA, October, pp 1218–1221
11. Hewlett-Packard Corporation, Intel Corporation, Microsoft Corporation, Phoenix Technologies Ltd., and Toshiba Corporation (2009) Advanced configuration and power interface specification, revision 4.0, June
12. Ginis G (2005) Low-power modes for ADSL2 and ADSL2+. White Paper, Broadband Communications Group, Texas Instruments, January
13. IEEE P802.3az (2010) Energy efficient Ethernet task force. http://grouper.ieee.org/groups/802/3/az
14. Chabarek J, Sommers J, Barford P, Estan C, Tsiang D, Wright S (2008) Power awareness in network design and routing. In: Proceedings of IEEE INFOCOM, Phoenix, USA, April, pp 457–465
15. Gupta M, Singh S (2003) Greening of the Internet. In: Proceedings of ACM SIGCOMM, Karlsruhe, Germany, August, pp 19–26
16. Mellah H, Sans'o B (2009) Review of facts, data and proposals for a greener Internet. In: Proceedings of ICST BROADNETS, Madrid, Spain, September, pp 1–5
17. Hwang C-H, Wu A (2000) A predictive system shutdown method for energy saving of event-driven. ACM Trans Des Autom Electron Syst 5(2):226–241
18. RedIRIS (2011) What is RedIRIS? http://www.rediris.es/rediris/index.html.en
19. Leinen S (2004) Evaluation of candidate protocols for IP flow information export (IPFIX). RFC 3955, October
20. Choi B-Y, Bhattacharrya S (2005) Observations on Cisco sampled Netflow. ACM SIGMETRICS Perform Eval Rev 33:18–23

21. Bellovin SM (2002) A technique for counting NATted hosts. In: Proceedings of ACM SIGCOMM workshop on Internet measurement, Marseille, France, November, pp 267–272
22. Postel J (ed) (1981) Internet control message protocol. RFC 792, USC/Information Sciences Institute, September
23. NMAP (2011) Network Mapper Security Scanner. http://nmap.org/
24. ARPing (2011) http://freshmeat.net/projects/arping/
25. Gouda MG, Liu AX (2005) A model of stateful firewalls and its properties. In: Proceedings of international conference on dependable systems and networks, Yokohama, Japan, June, pp 128–137
26. García-Dorado JL, Hernández JA, Aracil J, López de Vergara JE, Montserrat FJ, Robles E, de Miguel TP (2008) On the duration and spatial characteristics of Internet traffic measurement experiments. IEEE Commun Mag 46(11):148–155
27. Floyd S, Paxson V (2001) Difficulties in simulating the Internet. IEEE/ACM Trans Netw 9(4):392–403
28. Department of Energy and Environmental Protection Agency (2000) Carbon dioxide emissions from the generation of electric power in the United States, July
29. Bertoldi P, Atanasiu B (2007) Electricity consumption and efficiency trends in the enlarged European Union (status report 2006). UR 22753 EN, Institute for Environment and Sustainability, European Commission-JRC