Sustainable Computing: Informatics and Systems 10 (2016) 1–19

Contents lists available at ScienceDirect

Sustainable Computing: Informatics and Systems

journal homepage: www.elsevier.com/locate/suscom

Using Genetic Algorithms to optimise dynamic power saving in communication links subject to quality of service requirements

Huseyin Abaci [a,*], Gerard Parr [b], Sally McClean [b], Adrian Moore [b], Louise Krug [c]

[a] Adnan Menderes University, Computer Engineering, Central Campus, Aydın 09010, Turkey
[b] University of Ulster, School of Computing & Engineering, Coleraine BT52 1SA, Northern Ireland, UK
[c] BT Group, Adastral Park, Martlesham, Ipswich IP5 3RE, UK

ARTICLE INFO

Article history:
Received 4 September 2014
Received in revised form
17 November 2015
Accepted 8 January 2016
Available online 14 January 2016

Keywords:
Genetic Algorithm
Network performance
Power saving
Slowing Mechanism

ABSTRACT

Network devices, meeting increasing workload demand, are not efficiently Power-Workload Proportionate and consume a considerable amount of power even when the workload (utilisation) is low. This work proposes a novel Slowing Mechanism (SM) that provides Power Workload Proportionality for a wired network equipment to reduce power consumption. The Slowing will be achieved by adjusting the Operational Rate (OPR) of components according to traffic load. To meet applications' (VoIP, Data and Video) performance requirements, a Safety Gap (SG) is proposed in the Slowing Mechanism. Many parameters need to be carefully set for performance requirements within Slowing Mechanism. A Genetic Algorithm (GA) optimisation dynamically set to respond to the variable incoming traffic pattern determines these parameters. Thus, this work is a GA and the Slowing Mechanism integration to provide an insight into how GA optimisation can be employed in a network environment, and to optimise parameters in real-time.

The results demonstrate that a considerable amount of saving is achievable. With the default hardware configuration, the SM optimises the parameters and offers a saving of over 60% for typical stable traffic, with acceptable packet delay and no packet loss. This saving is reduced to 17% saving for a bursty traffic pattern with acceptable performance degradation.

1. Introduction

THE capacity of Information and Communication Technology (ICT) resources such as Integrated Circuits (IC) and mass storage devices doubles approximately every 18 months, as dictated by Moore's Law [1,2]. The associated improvement allows and encourages the developers deploying the ICT services to offer better facilities and services. Examples of the types of services provided include bandwidth intensive Video-on-Demand (VoD), Voice over Internet Protocol (VoIP), cloud computing, online multi-player gaming and file sharing. Previous estimates show that as a consequence of increasing Internet scale, ICT consumes 2–3% of the total global energy consumption [3,4]. Experts estimate that Internet Protocol (IP) traffic will continue to grow at 43% per annum, doubling every 1.4 years [5,6].

The pervasive computing, especially, has created increasing pressure on the network infrastructure due to the requirement for high volume data transmission between remote ends. This trend has already motivated providers to employ high capability power hungry network devices within their network, to enable their ability to scale to support this increase. Network devices, meeting increasing workload demand, are not efficiently Power-Workload Proportionate and consume a considerable amount of power even when the workload (utilisation) is low [7–9]. The results in [8] show that there is no wide application of Power Workload Proportionality in general switching and routing network equipment. While network devices have been used for energy reduction in different areas such as transportation, production and services, self-energy efficiency of network devices have been neglected. In consideration of all the above, urgent power-workload optimisation of the network infrastructure is required in order to establish more energy efficient networks by employing state-of-the-art power saving technologies [10–12].

This work proposes a novel Slowing Mechanism (SM) that provides Power Workload Proportionality for a wired communication link of the network equipment to reduce power consumption. The Slowing will be achieved by a Dynamic Voltage–Frequency Scaling

* Corresponding author. Tel.: +02562182000; fax: +90 2562136686.
E-mail addresses: huseyin.abaci@adu.edu.tr (H. Abaci), gp.parr@ulster.ac.uk (G. Parr), si.mcclean@ulster.ac.uk (S. McClean), aa.moore@ulster.ac.uk (A. Moore), louise.krug@bt.com (L. Krug).

http://dx.doi.org/10.1016/j.suscom.2016.01.002
2210-5379/

(DVFS) technique that increases or decreases voltage–frequency to adjust the Operational Rate (OPR). The OPR is the forwarding packets/bits rate of a link adjusted according to the workload. However, performance and power are trade-offs, and the Slowing power saving technique comes at the cost of network performance degradation. To meet applications' (VoIP, Data, and Video) performance requirements, a Safety Gap (SG) is proposed in the SM. Many parameters need to be carefully set for performance requirements within SM. These parameters are determined by a Genetic Algorithm (GA) optimisation tool dynamically set to respond to the variable incoming traffic pattern. Thus, this work is a prototype of GA and the Slowing Mechanism integration to provide an insight into how GA optimisation can be employed in a network environment, and to optimise parameters in real-time.

Moreover, performance degradation of the SM was investigated using a testbed by considering the performance metrics of the offered bandwidth, packet delay and packet loss. Based on the results collected, a trend line is identified that represents the correlation between the SG and the packet delay. This trend was then formulised to propose a cubic polynomial equation.

The results demonstrate that a considerable amount of saving is achievable. With the default hardware configuration, the SM optimises the parameters and offers a saving of over 60% for typical stable traffic, with acceptable packet delay and no packet loss. This saving is reduced to 17.5% saving for a bursty traffic pattern with acceptable performance degradation.

The structure of this paper is as follows: Section 2 reviews the background knowledge and related works of slowing saving technique, Section 3 presents Quality of Service (QoS) and algorithm of SM, Section 4 investigates QoS and performance degradation of SM, Section 5 explains the simulation methodology and SM's equations which are used in the Simulation, Section 6 selects well performed GA options for the SM and analyses the GA performance, Section 7 presents simulation results followed by conclusion and future direction of this work.

2. Background and related works

Slowing is achievable by sufficiently slowing down the OPR of the entire IC, subcomponents of the IC and the communication links according to load by voltage and/or frequency reduction. However, making network component go into less power consuming state is not an easy task. Performance and power are trade-offs and many parameters need to be carefully set for performance requirements within slowing. Due to frequency synchronisation between subcomponents, there remains a short operation cut, consequently a small amount of buffering and delay is possible before and during the state transition change. This can be perceived as a delay for real-time applications such as VoIP, IPTV at the receiver and this delay should not compromise the QoS metrics. For example, a metro/core domain contains high capability metro/core routers that aggregate link's traffic, connect networks, function as an interpreter as they look at addresses and pass packets to their destinations. These routers should remain up and running at all times, due to continuous heavy traffic on the links. Any shut down for even a millisecond may cause delay, extra buffering and significant packet loss at the busy link which would cause deterioration in the performance. The network devices should perform a state transaction quickly and eliminate unnecessary transitions.

Moreover, the load pattern changes by time and the spatial location of the device, and it is difficult to predict the exact time of the bursts. Accurate load prediction is essential with lowering the bitrates (slowing) and it is investigated by Observation Interval (OI). The OI is a number of samples that are collected to gather information about average utilisation, burstiness of traffic and number of packets/bits arrival by investigating the traffic pattern history in order to predict future load [63]. If the traffic pattern varies significantly, a larger OI may provide a better perspective to gather more accurate characteristics of the pattern than a shorter OI [13]. On the other hand, the larger OI may result less power saving. Therefore, careful selection of OI and the number of operation rates switched between, which will result in better performance-power trade-offs.

For a metro-core router: If it is not an extreme situation, such as link failure, then there are no extreme spikes in the traffic flow due to high number of link aggregation. However, the traffic pattern fluctuates and sudden spikes do occasionally occur at the metro-core router and usually at the Customer Premises Equipment (CPE) domain. For this reason, SG is an important component and integrated into the SM. The SG acts as a safety cushion for our performance, compensates for sudden spikes, and prevents frequent rate fluctuations. However, setting a boundary for this gap requires attention: If the safety gap is too small, it will degrade the performance e.g. higher buffering and longer packet delay, but establish higher energy savings. Conversely, if the gap is too large, performance will not be degraded but energy savings will be small. The OPR is the forwarding rate of the link and this is determined by the operating frequency. Higher frequency yields faster OPR (forwarding rate) and lower frequency yields slower OPR. To determine (predict) the required OPR without affecting QoS, the SM needs to observe the InT closely.

The SM has a relatively high number of variables/parameters and the complex dependence structure of these variables means slowing the problem for a wider and complex solution space (not smooth or unimodal). This results in a classic hill climbing or minimum seek algorithm which are inefficient. Therefore, a heuristic optimisation algorithm is required to find an optimal or near optimal solution by searching the solution space for optimising the SM variables and parameters according to variable traffic pattern.

The GA have been used in a wide variety of optimisation tasks [28–30] and provides fast evaluation (data training), that presents not a single solution (oppose to Neural Network, Bayesian Network) but a pool of possible good solutions. The GA is especially suitable for a problem even satisfied by a good solution (similar to SM in terms of power saving) than the optimum solution. The model is easily adapted to any problem if the fitness function is determined. In addition, the GA performs well if a problem has many dependent continuous variables as is the case with the SM, and provides good results with both high and medium dependencies [31].

Overall, the GA is adaptable, robust and flexible. It optimises with continuous or discrete variables, deals with a large number of variables and complex variable influence, resilient to being trapped in the local minima by mutation and crossover, provides quick convergence with the right parameters and does not have much mathematical requirement (less computational effort) [29,31,32]. Due to the outstanding advantages mentioned above and the comparative experiments performed in [66–70] the GA is an attractive/suitable and better tool for optimising the SM variables and the GA is selected to optimise variables in order to save power while providing acceptable performance.

2.1. Related works

The first introduction papers regarding link rate reduction was "Adaptive Link Rate" (slowing) proposed by Gunaratne and Christensen et al. [14–16]. The authors used policy-based mechanisms to determine when to adjust the bitrates for a LAN Ethernet and used a few recognised bitrates such as 10 Mb, 100 Mb and 1 Gb to switch between. The authors describe three policies (i) Duel-threshold policy, that base on queuing length and buffer occupancy. If buffer occupancy passes the high threshold level, higher bitrates set. If buffer occupancy is lower than the low threshold level mechanism