A Survey of Internet Energy Efficiency Metrics

Kerry Hinton[1] and Fatemeh Jalali[2]
[1]Centre for Energy Efficient Telecommunications, University of Melbourne, Parkville, Australia
[2]IBM Research, Melbourne, Australia

Keywords: Energy Consumption, Energy Efficiency, Communications Networks, Telecommunications Services, Internet Services.

Abstract: Several metrics have been widely applied to quantify the "energy efficiency" of the Internet and ICT. In this paper we analyse and compare these metrics when applied to telecommunication network equipment, networks and services. We show that different metrics can imply different, and possibly conflicting, strategies for improving energy efficiency. Some guidelines are suggested for the appropriate application of these metrics.

1 POWER & ENERGY MODELS

1.1 Equipment Power Model

The dependence of power consumption, $P(t)$, at time t, on traffic throughput, $C(t)$, for network equipment can be written in a generic "affine" form (Vishwanath, et al. 2014):

$$P(t) = P_{idle} + EC(t) = P_{idle} + \frac{(P_{max} - P_{idle})}{C_{max}} C(t) \quad (1)$$

Where P_{idle} is the power consumption with no throughput (i.e. $C(t) = 0$), C_{max} is the maximum throughput of the network element and P_{max} is the power consumption when $C(t) = C_{max}$. In (1) the linear slope $E = (P_{max} - P_{idle})/C_{max}$ has dimensions of energy per bit. We shall refer to this slope as the "incremental energy per bit".

1.2 Network Power Model

Using (1), the total network power, P_{Ntwk}, is the sum of the equipment power;

$$P_{Ntwk}(t) = \sum_{j=1}^{N_E} \left(P_{idle,j} + E_j C_j(t) \right) \quad (2)$$

where N_E is the number of network elements, $P_{idle,j}$, E_j and $C_j(t)$ are the idle power, incremental energy per bit and throughput of the j-th network element respectively.

We have $\sum_j C_j \geq C_{Ntwk}$, because most traffic flows will go through multiple network elements. If we identify all network traffic flows with a service, including network management, control and monitoring traffic, then we can write

$$C_{Ntwk}(t) = \sum_{k=1}^{N^{(S)}} C^{(k)}(t) \quad (3)$$

where $N^{(S)}$ is the number of services, k is the index for the service and $C^{(k)}$ is the traffic (bits/sec) for the k-th service. We also have for the traffic through the j-th network element:

$$C_j(t) = \sum_{k=1}^{N^{(S)}} C_j^{(k)}(t) = \sum_{k=1}^{N^{(S)}} \alpha_j^{(k)} C^{(k)}(t) \quad (4)$$

where $C_j^{(k)} = \alpha_j^{(k)} C^{(k)}$ is the traffic through the j-th network element due to service k. ($\alpha_j^{(k)}$ is the proportion of service traffic $C^{(k)}$ that propagates through network element j.) Note that for a service k, we also have $C^{(k)} \leq \sum_j C_j^{(k)}$.

1.3 Service Power Model

The "fundamental unit" of the Internet (as far as its end users are concerned) is "service" (e.g. SaaS, IaaS, Google Docs, Dropbox, etc). There has been a growing interest in the power consumption of services over recent years (Chan, C., et al., 2012).

From (1) we see the power consumption of a network element has an idle power component, P_{idle}, Today, most wireline network equipment has $P_{idle} \geq$

Hinton, K. and Jalali, F.
A Survey of Internet Energy Efficiency Metrics.
In *Proceedings of the 5th International Conference on Smart Cities and Green ICT Systems (SMARTGREENS 2016)*, pages 243-251
ISBN: 978-989-758-184-7

$0.8P_{max}$ (Vishwanath, A., et al. 2014). For much of the "Layer 0" and "Layer 1" equipment $P_{idle} = P_{max}$. To allocate power to services, we need an approach for allocating a proportion of P_{idle} to each service traffic flow propagating through a network element. If $P_{idle,j}^{(k)}(t)$ is the idle power allocated to service k in network element j, we expect the sum over all services through that element to satisfy

$$P_{idle,j} = \sum_{k=1} P_{idle,j}^{(k)}(t) \tag{5}$$

One way to fulfil this requirement, is to apply the same linear proportionality rule to the idle power as found for the incremental power in (1). That is, we set $P_{idle,j}^{(k)}(t) \propto C_j^{(k)}(t)$. Using (5), $P_{idle,j}^{(k)}(t)$ has the form:

$$P_{idle,j}^{(k)}(t) = P_{idle,j}\frac{C_j^{(k)}(t)}{C_j(t)} \tag{6}$$

With this rule, the overall power consumption of the k-th service, $P^{(k)}(t)$ provided by the network is

$$P^{(k)}(t) = \sum_{j=1}^{N_E}\left(\frac{P_{idle,j}}{C_j(t)} + E_j\right)C_j^{(k)}(t) \tag{7}$$

It is important to note that this is not the only approach available. For example, we could set $P_{idle,j}^{(k)}(t) = P_{idle,j}/N_j^{(srv)}(t)$ where $N_j^{(srv)}(t)$ is the number of services through network element j at time t. A disadvantage of this approach is that $N_j^{(srv)}(t)$ can be awkwardly large for core network equipment.

2 EFFICIENCY METRICS

The ITU has described an energy-efficiency metric in ITU-T Rec. L.1330 as (ITU-T 2012(a)):

"The energy efficiency metric is typically defined as the ratio between the functional unit and the energy necessary to deliver the functional unit."

This definition results in a metric with units "bits/Joule". ITU-T Rec. L.1330 also recognises that "The inverse metric, energy divided by functional unit, could be used as an alternative."

We shall focus on "energy per bit" metrics.

2.1 Standardised Metrics

Energy efficiency metrics currently used in standards documents are based on the ratio of power to traffic for a number, M, of pre-defined load levels of the equipment (Minoli, D., 2011). For example, the ECR-VL is defined by the ratio,

$$ECR = \sum_{m=1}^{M}(a_m \times P_m) \Big/ \sum_{m=1}^{M}(a_m \times C_m) \tag{8}$$

Where $\sum_m a_m = 1$, P_m and C_m are the power and the pre-defined loads indexed by m. The values of P_m and C_m are specified in the definition and depend upon the type of network element.

For a network element, placing (1) into (8) gives

$$ECR = \frac{P_{idle}}{C_{max}\sum_m a_m b_m} + E \tag{9}$$

Where $b_m = C_m/C_{max}$. Example values for b_m are: $b_1 = 1$, $b_2 = 0.5$, $b_3 = 0.3$, $b_4 = 0.1$ and $b_5 = 0$ with corresponding weights $a_1 = 0.1$, $a_2 = 0.5$, $a_3 = 0.3$, $a_4 = 0$ and $a_5 = 0.1$ (Minoli, D., 2011).

There are several problems with this and similar metrics (TEER, EER and TEEER) (Minoli, D., 2011). First, although these definitions include averages over loads, they effectively correspond to a single load; therefore the value does not incorporate the impact of traffic variation over the diurnal cycle.

Another is seen by considering two routers with the (approximately) same C_{max} but different values for P_{idle} and E. Let the values for the routers be $P_{idle,1}$, E_1 and $P_{idle,2}$ E_2 respectively and $P_{idle,2} = xP_{idle,1}$ where x is constant. Then we get the same ECR value for both routers provided:

$$E_2 = E_1 + \frac{P_{idle,1}}{C_{max}\sum_m a_m b_m}(1-x) \tag{10}$$

If we set $x = 0$ router 2 is load proportional (i.e. $P_{idle,2} = 0$) which is viewed as desirably energy efficient. However it still has the same ECR as router 1 which may have large idle power.

Using the value for x given by

$$x = 1 + \left(\frac{E_1 C_{max}}{P_{idle,1}}\right)\sum_m a_m b_m \tag{11}$$

in (10) gives $E_2 = 0$ which is often considered energy inefficient but will still have the same ECR as router 1. Hence ECR type metrics do not reflect the idea of "energy efficiency" very well.

Finally, the metric defined in (8) does not lend itself to being applied to networks or services. It is not clear how to apply this metric to a collection of interconnected network elements. Even less apparent is how to apply this metric to service that may be one of many propagating through an element.

2.2 Defining Energy Efficiency

In this work we will study a range of metrics that have been proposed and applied (Schien, D, and Preist, C. 2014)(GreenTouch, 2015)(Baliga, J., et al. 2009). We will implement them in a manner that is applicable to network elements, networks and services. Simple energy/bit efficiency metrics that have been employed in the literature are:

a) "Instantaneous energy per bit" defined by the ratio of the instantaneous power to throughput:

$$^{1}H(t) = \frac{P(t)}{C(t)} \tag{12}$$

This metric has been adopted in a range of "bottom-up" metrics used to calculate the energy efficiency of the Internet at peak load (Baliga, J., et al. 2009) or at time t (Chiaraviglio, L., et al., 2009). When $P_{idle} \neq 0$, $^{1}H(t)$ will vary over the diurnal cycle.

b) "Energy per bit" defined by the ratio of total energy, $\mathcal{E}(T)$, consumed over duration T to total bit throughput. $\mathcal{B}(T)$, over duration T:

$$^{2}H(T) = \frac{\int_0^T P(t)\,dt}{\int_0^T C(t)\,dt} = \frac{\mathcal{E}(T)}{\mathcal{B}(T)} = \frac{\langle P \rangle_T}{\langle C \rangle_T} \tag{13}$$

In this equation the time integral is over duration T from a pre-determined origin time. The GreenTouch consortium uses $1/(^{2}H(T))$ with T = 1 year for the years 2010 and 2020 (GreenTouch, 2015). In (13) $\langle X \rangle_T = \int_T X(t)dt/T$.

Some "top-down" metrics use (13), with $\mathcal{E}(T)$ determined from information such as equipment deployment inventory data and energy consumption and $\mathcal{B}(T)$ is an assessment of the total network traffic (Schien, D, and Preist, C. 2014).

c) "Mean instantaneous energy per bit" was proposed in (ITU-T 2012(b)), although that document contains a mathematical error. The average the instantaneous metric over time duration T is defined by:

$$^{3}H(T) = \frac{1}{T}\int_0^T \frac{P(t)}{C(t)}\,dt = \left\langle \frac{P}{C} \right\rangle_T = \left\langle {}^{1}H(t) \right\rangle_T \tag{14}$$

Other metrics have been defined, for example mobile network researchers have used power per unit area (Tombaz, S. et al. 2013). Due to lack of space, in this work we will focus on the energy per bit metrics, ^{1}H, ^{2}H and ^{3}H, listed above.

2.3 Uses of Metrics

Metrics are most frequently used for improvement (i.e developing strategies to change the value of the metric for a system), benchmarking (comparing the value of a metric for the systems being benchmarked) or estimating energy consumption. When used for improvement, the choice of metric will directly impact the strategies adopted for "improvement". When used for benchmarking, the choice of metric will determine what we mean when we say one system is "better" than another. Therefore, the choice of metric is important.

3 DIURNAL CYCLES

Diurnal traffic cycles result from the fact that many users are typically "off-line" and "on-line" during common times over a 24 hour period. An example of a diurnal cycle for an Australian city (taken from an edge router traffic log) is shown by the solid line in Fig. 1. Also shown is a pure sinusoidal approximation (dashed line) of the 24 hour diurnal cycle. In general, traffic diurnal cycles can dramatically vary in shape; however they all have a cyclic profile.

For the purposes of comparing the general characteristics of these metrics we shall use a "first-order" sinusoidal approximation for the diurnal cycle of the k-th service's traffic flow (in bits/sec) of the form

$$C^{(k)}(t) = C_{mean}^{(k)} + \Delta C^{(k)} \cos\left(2\pi t/T - \phi^{(k)}\right) \tag{15}$$

with T = 24 hours. In (15) $C_{mean}^{(k)}$ is the mean traffic for the k-th service flow over duration T and $\Delta C^{(k)}$ the variation away from the mean for the k-th flow. The phase $\phi^{(k)}$ accounts for the fact that the diurnal cycles of the individual services may not be synchronized (i.e. different services will have a different time of peak traffic). Using (15) enables the calculation of closed forms for the metrics above.

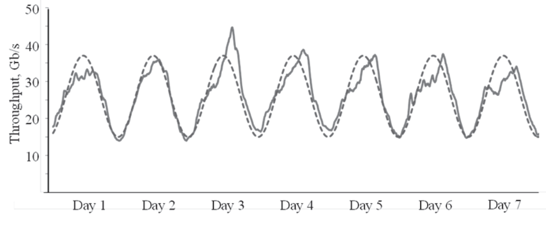

Figure 1: Example of a 7 day diurnal cycle from a commercially deployed router (solid line) and a first-order sinusoidal approximation (dashed line).

The form in (15) can be applied to equipment and networks. From (3) the time dependence of the total network traffic is the sum of all the service traffics:

$$\begin{aligned} C_{Ntwk}(t) &= C_{mean,Ntwk} + \Delta C_{Ntwk}\cos\left(\left(2\pi t/T\right)-\phi_{Ntwk}\right) \\ &= \sum_{k=1}^{N^{(S)}}\left(C_{mean}^{(k)} + \Delta C^{(k)}\cos\left(\left(\left(2\pi t\right)/T\right)-\phi^{(k)}\right)\right) \end{aligned} \quad (16)$$

Similarly, from (4) the total traffic for the j-th network element, $C_j(t)$, will have the form:

$$\begin{aligned} C_j(t) &= C_{mean,j} + \Delta C_j\cos\left(\left(2\pi t/T\right)-\phi_j\right) \\ &= \sum_{k=1}^{N^{(S)}}\left(C_{mean,j}^{(k)} + \Delta C_j^{(k)}\cos\left(\left(\left(2\pi t\right)/T\right)-\phi^{(k)}\right)\right) \\ &= \sum_{k=1}^{N^{(S)}}\alpha_j^{(k)}\left(C_{mean}^{(k)} + \Delta C^{(k)}\cos\left(\left(\left(2\pi t\right)/T\right)-\phi^{(k)}\right)\right) \end{aligned} \quad (17)$$

The ϕ terms correspond to the location of the peak load within the diurnal cycle relative to a fixed arbitrary origin. Setting the time origin to time t_0 (hours) then if the peak traffic occurs at time t_{peak}, we have $\phi = 2\pi\,(t_{peak} - t_0)/24$.

For a network or element in which the flows are synchronized (i.e. all flows have the same peak traffic time) we have

$$\begin{aligned} C_{mean,Ntwk} &= \sum_{k=1}^{N^{(S)}} C_{mean}^{(k)} \qquad C_{max,Ntwk} = \sum_{k=1}^{N^{(S)}} C_{max}^{(k)} \\ C_{mean,j} &= \sum_{k=1}^{N^{(S)}} C_{mean,j}^{(k)} \qquad C_{max,j} = \sum_{k=1}^{N^{(S)}} C_{max,j}^{(k)} \end{aligned} \quad (18)$$

However, with unsynchronized flows, we get

$$\begin{aligned} C_{Ntwk}(t) &= \sum_{k=1}^{N^{(S)}} C_{mean}^{(k)} + \\ &\left\{\sum_{k=1}^{N^{(S)}}\left(\Delta C^{(k)}\right)^2 + \sum_{k\neq l}^{N^{(S)}}\sum_{l}^{N^{(S)}}\Delta C^{(k)}\Delta C^{(l)}\cos\left(\phi^{(k)}-\phi^{(l)}\right)\right\}^{1/2} \\ &\times\cos\left(\left(2\pi t/T\right)-\phi_{Ntwk}\right) \end{aligned} \quad (19)$$

where

$$\phi_{Ntwk} = -\tan^{-1}\left(\sum_{k=1}^{N^{(S)}}\Delta C^{(k)}\sin\phi^{(k)} \Big/ \sum_{k=1}^{N_S}\Delta C^{(k)}\cos\phi^{(k)}\right) \quad (20)$$

A corresponding form can be written for $C_j(t)$.

Assuming the differences ($\phi^{(k)} - \phi^{(l)}$) is are not all zero, then comparing a network with many synchronized flows to the same network with many unsynchronized flows we find,

$$\left.\frac{\Delta C_{Ntwk}}{C_{mean,Ntwk}}\right|_{synch} > \left.\frac{\Delta C_{Ntwk}}{C_{mean,Ntwk}}\right|_{unsynch} \quad (21)$$

Taking this further, if the phases, $\phi^{(k)}$, are uniformly, randomly distributed

$$\left.\frac{\Delta C_{Ntwk}}{C_{mean,Ntwk}}\right|_{unsynch} \sim \frac{\left\langle\left(\Delta C^{(k)}\right)^2\right\rangle^{1/2}}{\sqrt{N^{(S)}}\left\langle C_{mean}^{(k)}\right\rangle} \quad (22)$$

From this we see that as the number of unsynchronised flows increases the network traffic maximum, given by $C_{mean,Ntwk} + \Delta C_{Ntwk}$, reduces to the mean, $C_{mean,Ntwk}$;

$$\begin{aligned} &\lim_{N^{(S)}\to\infty} C_{\max,Ntwk} = \\ &\lim_{N^{(S)}\to\infty} C_{mean,Ntwk}\left(1+\frac{\Delta C_{Ntwk}}{C_{mean,Ntwk}}\right) \sim C_{mean,Ntwk} \end{aligned} \quad (23)$$

This means the depth of the diurnal cycle reduces when increasingly many unsynchronised traffic flows are brought together. This result applies to any single network element, network facility or overall network that deals with many service flows.

The results in (21) (22) and (23) tell us that facilities dealing with highly synchronised traffic (such as serving only a local time-zone) are likely to experience a relatively deeper diurnal cycle than those dealing with unsynchronised traffic (such as traffic from geographically diverse regions around the globe).

Diurnal cycle depth plays an important role when improving energy efficiency, because networks are dimensioned to accommodate peak traffic (C_{max}). In legacy networks equipment remains fully energised 24/7, therefore dimensioning network for peak load means that during off-peak hours equipment is lowly utilised which is less energy inefficient (i.e. higher energy per bit) than at peak time.

In new generation networks, a widely proposed strategy is to implement low energy (sleep) states during "off-peak" times to improve energy efficiency (Mahadevan, P. et al., 2009). The depth of the diurnal cycle is important because it indicates how much equipment can be powered-down during off peak times (GreenTouch, 2015).

4 THE METRICS

4.1 Network Equipment

Without a loss of generality, we can drop the phase term when applying the diurnal cycle to the traffic through a network element. For the j-th network element, we have

$$
\begin{aligned}
{}^{1}H_j(t) &= \frac{P_{idle,j}}{C_{mean,j} + \Delta C_j \cos(2\pi t/T)} + E_j \\
{}^{2}H_j(T) &= \frac{P_{idle,j}}{C_{mean,j}} + E_j \\
{}^{3}H_j(T) &= \frac{P_{idle,j}}{C_{min,j}^{1/2} C_{max,j}^{1/2}} + E_j
\end{aligned}
\tag{24}
$$

where $C_{min,j} = C_{mean,j} - \Delta C_j$. To calculate ${}^{3}H_j(T)$ we have used Item 3.613.1 from (Gradshteyn, I. Ryzhik, I. 1980).

We note that $C_{mean} = \langle C \rangle_T$, therefore we could interpret the difference between ${}^{2}H(T)$ and ${}^{3}H(T)$ results from the former using the arithmetic mean of the traffic $C(t)$ whereas the latter uses the geometric mean (for sinusoidal traffic load).

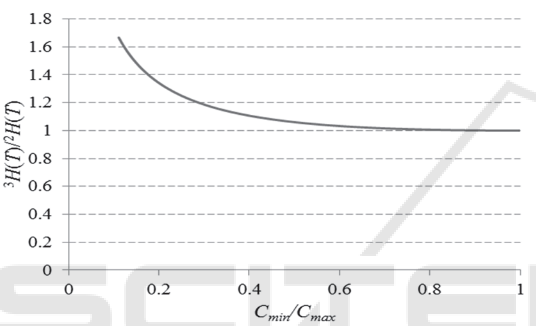

Figure 2: Plot of the ratio ${}^{3}H(T)/{}^{2}H(T)$ showing that the metric ${}^{3}H(T)$ reflects the impact of durations of low utilisation on network energy/bit metric for equipment in which idle power dominates.

Comparing ${}^{2}H_j(T)$ and ${}^{3}H_j(T)$ in (24) we note that ${}^{3}H_j(T)$ reflects the impact of traffic variation over a diurnal cycle where-as ${}^{2}H_j(T)$ does not. We see this by calculating the ratio ${}^{3}H_j(T)/{}^{2}H_j(T)$ over a range diurnal cycle depths (C_{min}/C_{max}) for the network element traffic as shown in Fig. 2. ${}^{2}H_j(T)$ is constant with respect to the ratio (C_{min}/C_{max}) where-as ${}^{3}H_j(T)$ exposes the impact of periods of low utilisation which correspond to low values of (C_{min}/C_{max}).

4.2 Networks

For a network using (2) and (3) we have

$$
{}^{1}H_{Ntwk}(t) = \frac{\sum_{j=1}^{N_E} \left(P_{idle,j} + E_j \left(C_{mean,j} + \Delta C_j \cos\left(\frac{2\pi t}{T} - \phi_j \right) \right) \right)}{C_{mean,Ntwk} + \Delta C_{Ntwk} \cos((2\pi t)/T)} \tag{25}
$$

$$
{}^{2}H_{Ntwk}(T) = \frac{\sum_{j=1}^{N_E} \left(P_{idle,j} + E_j C_{mean,j} \right)}{C_{mean,Ntwk}} \approx \frac{N_E \left(\langle P_{idle} \rangle_E + \langle E \rangle_E \langle C_{mean} \rangle_E \right)}{C_{mean,Ntwk}} \tag{26}
$$

$$
\begin{aligned}
{}^{3}H_{Ntwk}(T) = \frac{1}{C_{min,Ntwk}^{1/2} C_{max,Ntwk}^{1/2}} \sum_{j=1}^{N_E} \Bigg(& P_{idle,j} + \\
& E_j C_{mean,j} \left(1 - \frac{\Delta C_j}{C_{mean,j}} \frac{\left(C_{max,Ntwk}^{1/2} - C_{min,Ntwk}^{1/2} \right)}{\left(C_{max,Ntwk}^{1/2} + C_{min,Ntwk}^{1/2} \right)} \cos\phi_j \right) \Bigg)
\end{aligned}
\tag{27}
$$

The notation $\langle X \rangle_E$ represents an average over network elements, defined by

$$
\langle X \rangle_E = \frac{1}{N_E} \sum_{j=1}^{N_E} X_j \tag{28}
$$

The approximations in (26) and (27)is justified by the fact that E_j and C_j are independent random variables hence $\langle EC \rangle_E \approx \langle E \rangle_E \langle C \rangle_E$.

In ${}^{3}H_{Ntwk}(T)$, ϕ_j is measured relative to the network peak traffic time, that is $\phi_j = 2\pi\ (t_{peak,j} - t_{peak,Ntwk})/24$. To calculate (27) we have used Items 2.554.2 and 2.553.2 from (Gradshteyn, I. Ryzhik, I. 1980).

From (27) the factors that feed into ${}^{3}H_{Ntwk}(T)$ are the relative depths of the traffic diurnal cycles, $\Delta C/C_{mean}$, and the degree of synchronisation of the traffic flows, $\phi^{(k)}$ (see (17)). To acquire an appreciation of the impact of these parameters on the metrics, a mesh network simulation was constructed. The simulated network consisted of 50 inter-connected network elements ($N_E = 50$) each with a power profile given by (1) with a range of values for P_{idle} (randomly selected in the range 1kW to 1.5kW) and E (randomly selected in the range 0.5 nJ/bit to 2 nJ/bit). These values are typical of current generation router and switch technology (Van Heggegham, W., et al., 2012). The network carries 500 sinusoidal service flows ($N^{(S)} = 500$), with mean flow data rates randomly distributed over the range 0.5 Gbit/s $\leq C_{mean} \leq$ 2 Gbit/s. Each flow travels through 10 network elements randomly selected from the 50 elements in the simulation. No flow travels through the same element more than once. Although the simulation is a mesh network, the architecture is not a major influence because the overall power consumption is determined by the equipment along the service flow paths not the global network architecture.

The synchronisation of the flows is parametrised by quantity b with the phase of the flows chosen randomly over the range $-b\pi \leq \phi^{(k)} \leq b\pi$. For highly synchronised flows we set $b = 0.1$. For totally desynchronised flows we set $b = 1$. The simulation is run for values of b from 0.1to 1.0 in steps of 0.1.

To parametrise the diurnal cycle depth, $\Delta C/C_{mean}$ the flows are distributed randomly over a range $\Delta C/C_{mean}$ to $\Delta C/C_{mean} + 0.2$ with values of $\Delta C/C_{mean}$

from 0.1 to 0.8 in steps of 0.1.

For the case of synchronised flows and a deep diurnal cycle in the simulated network, the value of ${}^1H_{Ntwk}(t)$ can vary dramatically over the diurnal cycle. For the simulated network described above, at peak traffic time, we get ${}^1H_{Ntwk}(t_{peak}) \approx 66$ nJ/bit. At the time of minimum traffic (t_{trough}) ${}^1H_{Ntwk}(t_{trough}) \approx 910$ nJ/bit. Therefore using this metric requires careful consideration of the time at which it is measured. Measuring at peak traffic time will give a low estimate for typical energy per bit. From (25) and (26), for a totally desynchronised network we have ${}^1H_{Ntwk}(t) \approx {}^2H_{Ntwk}(T)$ for all t over the diurnal cycle. For the simulated network, this situation gives ${}^1H_{Ntwk}(t) \approx {}^2H_{Ntkw}(T) \approx 113$ nJ/bit.

The values of ${}^2H_{Ntwk}(T)$ and ${}^3H_{Ntwk}(T)$ (T = 24 hours) for ranges of cycle depth, $\Delta C/C_{mean}$, and synchronisation, b, are shown by the surface plot in Fig. 3. The top left region of the surface plots corresponds to highly synchronised service traffic flows with relatively deep diurnal cycles. We see that ${}^3H_{Ntwk}(T)$ reflects the impact of periods of low network utilisation that occur in networks with highly synchronised traffic and a deep diurnal cycle. In contrast ${}^2H_{Ntwk}(T) \approx {}^3H_{Ntwk}(T)$ for networks that are desynchronised or have shallow diurnal cycles.

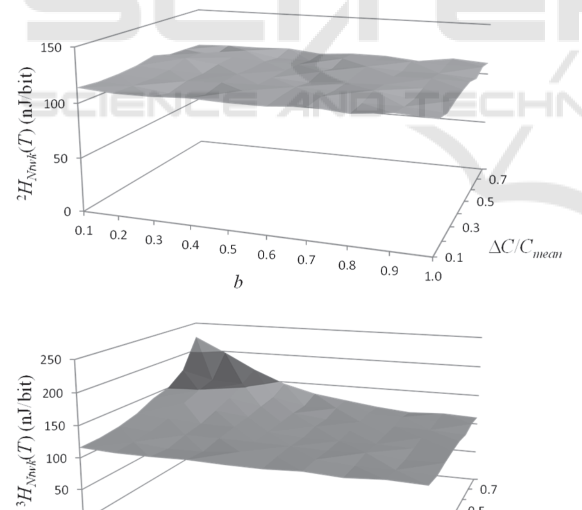

Figure 3: Surface plots of ${}^2H_{Ntwk}(T)$ and ${}^3H_{Ntwk}(T)$ for values of synchronicity parameter b and diurnal cycle depth $\Delta C/C_{mean}$.

4.3 Services

Using the equations above, for the k-th service

$$ {}^1H^{(k)}(t) = \sum_{j=1}^{N_E} \alpha_j^{(k)} \left(\frac{P_{idle,j}}{C_j(t)} + E_j \right) = \sum_{j=1}^{N_E} \alpha_j^{(k)1} H_j(t) \tag{29} $$

$$ \begin{aligned} {}^2H^{(k)}(T) = & \\ & \sum_{j=1}^{N_E} \frac{\alpha_j^{(k)} P_{idle,j}}{C_{min,j}^{1/2} C_{max,j}^{1/2}} \left(1 - \frac{\Delta C^{(k)}}{C_{mean}^{(k)}} \frac{\left(C_{max,j}^{1/2} - C_{min,j}^{1/2} \right)}{\left(C_{max,j}^{1/2} + C_{min,j}^{1/2} \right)} \cos \phi_j^{(k)} \right) + \\ & + \sum_{j=1}^{N_E} \alpha_j^{(k)} E_j \end{aligned} \tag{30} $$

$$ {}^3H^{(k)}(T) = \sum_{j=1}^{N_E} \alpha_j^{(k)} \left(\frac{P_{idle,j}}{C_{min,j}^{1/2} C_{max,j}^{1/2}} + E_j \right) = \sum_{j=1}^{N_E} \alpha_j^{(k)3} H_j(T) \tag{31} $$

In (30) $\phi_j^{(k)}$ is the offset between the peak traffic time of the j-th network and the peak traffic time of the k-th service.

If all the services in the network are similarly synchronised (i.e. no service is significantly out of synch with all the other services) we get for all k;

$$ {}^2H^{(k)}(T) \approx \sum_{j=1}^{N_E} \frac{\alpha_j^{(k)} P_{idle,j}}{C_{mean,j}} + \alpha_j^{(k)} E_j \tag{32} $$

To graphically display the dependence of ${}^2H^{(k)}(T)$ and ${}^3H^{(k)}(T)$ on the parameters $\Delta C/C_{mean}$ and b, we average over the k-index; i.e. over the services giving;

$$ \left\langle {}^XH^{(k)}(T) \right\rangle^{(S)} = \sum_{k=1}^{N^{(S)}} {}^XH^{(k)}(T) \Big/ N^{(S)} \tag{33} $$

where X is 2 or 3.

Plotting the simulation results for $\langle {}^2H^{(k)}(T) \rangle^{(S)}$ and $\langle {}^3H^{(k)}(T) \rangle^{(S)}$, we get identical surface plots as in fig. 3. That is, for the simulation scenario ${}^2H_{Ntwk}(T) = \langle {}^2H^{(k)}(T) \rangle^{(S)}$ and ${}^3H_{Ntwk}(T) = \langle {}^3H^{(k)}(T) \rangle^{(S)}$. This can be shown to hold in general provided all of $C_{mean,j}$ fall within a limited range of values.

It is important to note that ${}^XH^{(k)}(T) \sim {}^XH_{Ntwk}(T)$ only applies when all the services are all similarly synchronised such that the phase, $\phi^{(k)}$, of each service is within a given range (parametrised by b) of all other services.

For a service that is significantly out of synchronisation with the other services, the value of ${}^2H^{(k)}(T)$ is significantly greater. For example, consider a service k for which $\phi^{(k)}$ well away from zero and $\phi^{(l)} \sim 0$ for all $l \neq k$. In this case the network is highly synchronised with only flow k well out of synch. In this situation, $C_{max,j} >> C_{min,j}$ for all j. Using (30) we get

$$ \begin{aligned} {}^2H^{(k)}(T) & \\ & \approx \sum_{j=1}^{N_E} \frac{\alpha_j^{(k)} P_{idle,j}}{C_{min,j}^{1/2} C_{max,j}^{1/2}} \left(1 + \frac{\Delta C^{(k)}}{C_{mean}^{(k)}} \right) + \sum_{j=1}^{N_E} \alpha_j^{(k)} E_j \qquad \phi^{(k)} \approx \pi \\ & \approx \sum_{j=1}^{N_E} \frac{\alpha_j^{(k)} P_{idle,j}}{C_{min,j}^{1/2} C_{max,j}^{1/2}} + \sum_{j=1}^{N_E} \alpha_j^{(k)} E_j \qquad \phi^{(k)} \approx \pi/2 \end{aligned} \tag{34} $$

Fig. 4 is a surface plot of ${}^2H^{(k)}(T)$ for a service with $\phi^{(k)} = \pi$ for all values of b. We see that an out-of-synch service has much higher energy per bit than the other, (in-synch) services when the network is highly synchronised. As the degree of synchronisation reduces or the diurnal cycle depth reduces, ${}^2H^{(k)}(T)$ reduces to that of the other services.

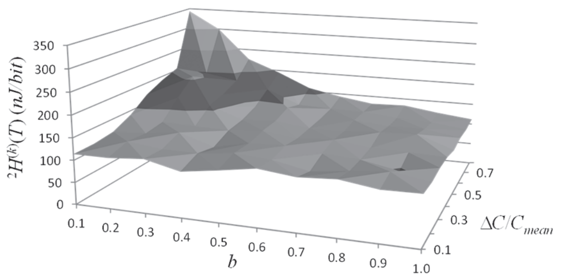

Figure 4: Surface plot for ${}^2H^{(k)}(T)$ for a service (π) out of phase with all other services in the network. Low values of b correspond to all other flows across the network being highly synchronised.

Comparing Fig. 3(a) and Fig. 4, we see that service operators wishing to minimise the energy per bit of their service will want to avoid being significantly out-of-synch with the majority of services. This will lead to service providers trying to synchronise their services with everyone else. This, in turn, will lead to deeper diurnal cycles and the resulting over-dimensioning of the network, mentioned in Section 4, and consequential increase in energy consumption.

5 APPLICATIONS

The expressions for the H metrics above are all based on a pure sinusoidal diurnal cycle. In real networks the diurnal cycle is not a pure sinusoid. However, generalising these metrics to arbitrary diurnal cycle profiles is relatively simple. For ${}^1H(t)$, we just replace the sinusoid with the actual diurnal cycle from collected traffic data. Because ${}^2H_j(T)$ and ${}^2H_{Ntwk}(T)$ only involve C_{mean}, these are directly applicable to any diurnal cycle profile.

The ${}^2H^{(k)}$(t) and ${}^3H$ metrics involve quantities C_{max}, C_{min}, ΔC and ϕ. To generalise these metrics we replace these values, in the metric definitions, with their means over multiple diurnal cycles: $\langle C_{max}\rangle_D$, $\langle C_{min}\rangle_D$, $\langle \Delta C\rangle_D$ and $\langle \phi\rangle_D$ which can be extracted from traffic data collected over multiple days. Where a quantity is raised to a power, a, we replace X^a by $(\langle X\rangle_D)^a$. Our discussion from now on can be applied to these generalised forms.

As discussed above, metrics ${}^1H(t)$ and ${}^2H(T)$ are already widely used where-as the ${}^3H(T)$ metric is not. The advantage provided by the ${}^3H(T)$ metric is that it quantifies the impact of the shape of the diurnal cycle and its relationship to other traffic flows (via C_{max}, C_{min} and ϕ). This enables us to quantify the impact of changing traffic profiles on energy efficiency of networks and services.

Although the energy efficiency metrics have primarily been created to provide a quantitative measure of "energy efficiency" (ITU-T 2012 (a))(Coroama, V. Hilty, L. 2014), they have been also used to estimate the power consumption of equipment, networks and services (Baliga, J., et al. 2009)(Van Heggegham, W., et al., 2012)(Vishwanath, A., et al. 2015). We will now consider some issues with these applications

5.1 Deployed Networks

The application of the metrics above in real networks can be very problematic due to unavailability of or difficulty in attainting the required data. In particular, evaluating these metrics for a network or service may require collection of a significant amount of data not readily available. Therefore approximations for the metrics can make evaluation easier, although possibly at the cost of reduced accuracy. Also, the inter-relationships between the metrics may allow the data collected for one metric to be used to evaluate another.

Using (26) we can show that

$$ {}^2H_{Ntwk} \approx \left(\left\langle N_{hops}\right\rangle + 1\right)\left\{\frac{\left\langle P_{idle}\right\rangle}{\left\langle C_E\right\rangle} + \left\langle E_E\right\rangle\right\} \tag{35} $$

where $\langle N_{hops}\rangle$ is the mean number of hops for service traffic across the network. This form aligns with the expressions for edge and core network energy efficiency in (Baliga, J., et al. 2009)(Van Heggegham, W., et al., 2012).

As discussed above, the simulation results show the ${}^2H$ metric for a network is approximately equal to the mean ${}^2H$ metric across the services, that is: ${}^2H_{Ntwk}(T) \approx \langle {}^2H^{(k)}(T)\rangle^{(S)}$. The results also show the variance $\sigma^{(S)}$ of the services satisfies $\sigma^{(S)}({}^XH^{(k)}(T)) < 0.1\ \langle {}^XH^{(k)}(T)\rangle^{(S)}$. This means that, to a first order approximation, provided all the services in the network are roughly synchronised to the same degree (i.e. no services are significantly out of synchronisation with the other services), we have

$$ {}^2H^{(k)}(T) \approx {}^2H_{Ntwk}(T) \tag{36} $$

for most of the services transported by the network. Similar results hold for ${}^3H^{(k)}(T)$ and ${}^3H_{Ntwk}(T)$. In this case, we can use the ${}^2H$ metric of a service to estimate the ${}^2H$ metric of a network or vice versa.

5.2 Estimating Power Consumption

Using energy efficiency to estimate power or energy consumption is based on the principle that the power, P, consumed by a network element, network or service with energy efficiency H joules/bit with traffic load C bit/sec is given by $P = HC$ (Baliga, J., et al. 2009)(Van Heggegham, W., et al., 2012). The energy consumption is given by $Q = HB$ where B is the number of bits transferred (Vishwanath, A., et al. 2015).

Although this appears to be intuitive, as we have seen above, there are multiple choices for evaluating H. Many authors have used the definition $H' = P_{max}/C_{max}$ where the values of P_{max} and C_{max} are based on data provided in equipment specification sheets or some form of measurement (Baliga, J., et al. 2009)(Van Heggegham, W., et al., 2012).

In some cases the utilisation U has been included to give $H'' = P_{max}/UC_{max}$ (Makkes, M., et al., 2013) where the "utilisation", $U = C(t)/C_{max}$.

In most cases, H is used to calculate the power or energy consumption of a service or user, based upon a data rate for the service or user, $C^{(k)}(t)$. Therefore, the appropriate equation is (7). Noting that $P_{max} = P_{idle} + EC_{max}$, we get

$$P^{(k)}(t) = \sum_j \left(\frac{P_{max,j}}{C_{max,j}} + \frac{P_{idle,j}}{C_{max,j}} \left(\frac{1}{U_j(t)} - 1 \right) \right) C^{(k)}(t) \qquad (37)$$

where the j-sum is over equipment along the path of the service data.

We see that for load proportional equipment ($P_{idle} \approx 0$), H' is appropriate and for constant power equipment ($P_{idle} \approx P_{max}$) the H'' is appropriate.

Comparing (13) and (14), ${}^2H(T)$ is more appropriate to calculate the power or energy consumption of a service because it has the form $\langle P\rangle_T/\langle C\rangle_T$ and a service is typically parametrised with $\langle C\rangle_T$ or $\mathcal{B}(T)$. Provided the conditions for (36) to hold are satisfied (see Sec. 8.1) we will have

$$\left\langle P^{(k)} \right\rangle_T = {}^2H_{Ntwk}(T) \left\langle C^{(k)} \right\rangle_T, \quad \mathcal{E}^{(k)}(T) = {}^2H_{Ntwk}(T)\mathcal{B}^{(k)}(T) \qquad (38)$$

This approach has been widely used (Vishwanath, A., et al. 2015) to estimate power or energy consumption of a variety of Internet services.

As shown in Sec. 7, (38) is only accurate if the service in question is not out-of-synch with the other network flows. Therefore using (38) to estimate the energy consumption of out-of-synch services (such as off-peak data transfer services) is inappropriate. This will also apply to services that travel through time-zones out-of-synch with their originator.

6 CONCLUSIONS

As a measure for energy efficiency, we have shown that, even for a given network and the values of metrics ${}^1H$, ${}^2H$ and ${}^3H$ can be significantly different. For example, in a network with somewhat synchronised traffic, the using the ${}^1H$ metric at peak traffic hour will give a very different value to the ${}^2H$ metric. Therefore, comparing these metrics can be problematic (Coroama, V. Hilty, L. 2014)(Schien, D, Preist, C. 2014).

When used to benchmark or improve energy efficiency, we see that desynchronising traffic flows reduces the ${}^3H$ metrics. Therefore, according to these metrics we can improve the energy efficiency of a network element by desynchronising its traffic flows. In contrast desynchronising flows has no impact on ${}^2H_j(T)$ and ${}^2H_{Ntwk}(T)$.

On the other hand, ${}^2H^{(k)}(T)$ indicates service providers should endeavour to synchronise their service flows with any oscillation in the diurnal cycle. If all service providers do this, the diurnal cycle will increase in depth which will impact network dimensioning. Hence the choice of metric influences strategy choice for improvement and even using the same metric in different situations may lead to different (and possibly conflicting) strategies.

We have also shown that, in certain circumstances, energy efficiency metrics can be used to estimate power or energy consumption. However, this must be done with care and particular note of how synchronised the service traffic is with other traffic in the network.

As the energy consumption of the Internet and ICT increases over the coming years, energy efficiency metrics will play an important role in mitigating this increase. In this paper we have summarised some of the subtleties that need to be considered in the application of these metrics.

REFERENCES

Baliga, J., et al. 2009, "Energy Consumption in Optical IP Networks", *IEEE Jour Lightwave Technol, Vol.27*, No.13, p.2391.

Chan, C., et al., 2012 “Methodologies for assessing the use-phase power consumption and greenhouse gas emissions of telecommunications network services” *Environ. Sci. Technol., Vol.47*, No.1, p.485.

Chiaraviglio, L., et al., 2009 “Energy-Aware Backbone Networks: A Case Study*”, ICC Workshop*.

Coroama, V. Hilty, L. 2014, “Assessing Internet energy intensity: A review of methods and results”, *Environmental Impact Assessment Review, Vol. 45*, p.63.

Gradshteyn, I. Ryzhik, I. 1980, “Table of Integreals, Series and Products”, Academic Press.

GreenTouch, 2015 “Reducing the Energy Consumption in Networks by up to 98% by 2020” *White paper, June*.

ITU-T 2012 (a), L.1310 “Energy efficiency metrics and measurement methods for telecommunications equipment” *ITU-T 11/2012*.

ITU-T 2012(b) Y.3022 (08/2012), “Measuring Energy in Networks”

Mahadevan, P. et al., 2009 “Energy aware network operations”, IEEE INFOCOM’09, p.25.

Makkes, M., et al., 2013 “A decision framework for placement of applications in clouds that minimizes their carbon footprint” *Jour. Cloud Computing, Vol.2*, No. 21.

Minoli, D., 2011 “Designing Green Networks and Network Operations” CRC Press.

Schien, D, Preist, C. 2014 “Approaches to Energy Intensity of the Internet”, *IEEE Communications Magazine, Nov*. p.130.

Tombaz, S. et al. 2013, “Energy Efficiency Assessment of Wireless Access Networks Utilizing Indoor Base Stations”, *PIMRC 2013*, p. 3105.

Van Heggegham, W., et al., 2012, “Power consumption modelling in optical multilayer networks”, *Photonic Network Communications, Vol.24, No.2, p.86.*

Vishwanath, A., et al. 2014 “Modelling Router Energy Consumption in High-Capacity Routers and Switches”, *IEEE JSAC., Vol.32, No.8*, p.*1524.*

Vishwanath, A., et al. 2015 “Energy Consumption Comparison of Interactive Cloud-Based and Local Applications”, *IEEE JSAC, Vol.33, No.4, p.616.*