



MAXIMIZE THE EFFICIENCY OF YOUR SERVER FARM

HOW TO REDUCE THE LEVEL OF ENERGY USED BY YOUR SERVERS WHEN
THEY'RE NOT BEING PRODUCTIVE

Mark Blackburn, Andy Dominey – 1E

ABSTRACT: THIS DOCUMENT SHOWS THE ISSUES THAT DATA CENTER MANAGERS FACE WHEN ATTEMPTING TO DISCOVER WHAT THE SERVERS IN THEIR CARE ARE DOING, AND HOW NIGHTWATCHMAN SERVER EDITION CAN EXPOSE THOSE SERVERS (BOTH PHYSICAL AND VIRTUAL) THAT ARE NOT PERFORMING USEFUL WORK SO THAT THEY CAN BE DECOMMISSIONED.

All rights reserved. No part of this document shall be reproduced, stored in a retrieval system, or transmitted by any means, electronic, mechanical, photocopying, recording, or otherwise, without permission from 1E. No patent liability is assumed with respect to the use of the information contained herein. Although every precaution has been taken in the preparation of this document, 1E and the authors assume no responsibility for errors or omissions. Neither is liability assumed for damages resulting from the information contained herein.

The 1E name is a registered trademark of 1E in the UK, US and EC. The 1E logo is a registered trademark of 1E in the UK, EC and under the Madrid protocol. NightWatchman is a registered trademark in the US and EU.



The physical problem

The commoditization of server hardware and the constant expansion in business application development over the last five years has led to a situation where almost every new application provisioned into the data center ended up with its own set of servers.

The hardware has become relatively cheap making it easier to obtain budget allowing each department or project to afford to have its own servers, removing the need to ask for shared computing resource and the delays and political maneuvering which that could cause. It also reduces risk to run each application on its own set of servers, since then there is no unforeseen interaction between applications on the same server to worry about, making the solution less likely to go wrong and therefore easier to support.

However the constant addition of new servers using more powerful components in denser enclosures (especially blade servers) means that power and cooling density requirements are now higher than most data center designed capacities, and hence there is a major drive towards data center efficiency and virtualization as means of combating these issues.

This rapid increase in the numbers of servers in a data center brings with it an administration and support overhead, and it becomes increasingly difficult to keep tabs on what every server is doing and if it is still required, with the result that servers fall out of use for various reasons (the application itself is no longer required, mergers and acquisitions/redundancy removes the need for an application, the project for which the servers were bought gets cancelled, the business function gets replaced by new applications, etc.).

When this happens the surplus servers are not removed from the environment or re-utilized elsewhere since the IT department is not aware of the fact that the servers are no longer required, and indeed our Server Energy and Efficiency Report found that almost half of the respondents of the survey believe that more than 15% of the servers that they are responsible for are doing no productive work.

The virtual problem

The use of server virtualization solutions, which can alleviate the pressures caused by the rapid rise in numbers of servers, can ironically also cause another more insidious version of the same issue to occur – virtual sprawl.

It is now so easy to rapidly provision a new virtual server, normally without the need for further capital expenditure, that very many more virtual servers are created than physical servers would have been without virtualization. People can spawn a new virtual server to test out an idea or try an application before committing to roll it out and, unless very tight controls are maintained, these virtual servers can be left behind unused when the testing has finished. Although they take up no physical space, and do not cause the underlying host hardware to draw any more power or require extra cooling, they tie up valuable disk, memory and CPU resources that could be more usefully deployed for productive work.

Without a mechanism for determining which servers are performing useful work it is very easy to end up with servers powered on but doing nothing useful, or to virtualize unused servers as part of a virtualization exercise, which is then compounded by the problem of virtual sprawl.

It is understandable that this situation occurs as it is actually quite difficult to determine if a server is no longer in use. Most centralized monitoring solutions are only designed to look at very high level performance metrics (Total CPU utilization, total I/O etc.) and these figures can be entirely misleading due to the fact that most well run data centers will use many different systems maintenance tools on their servers, and this maintenance activity masks the fact that the server may not be being used at all.

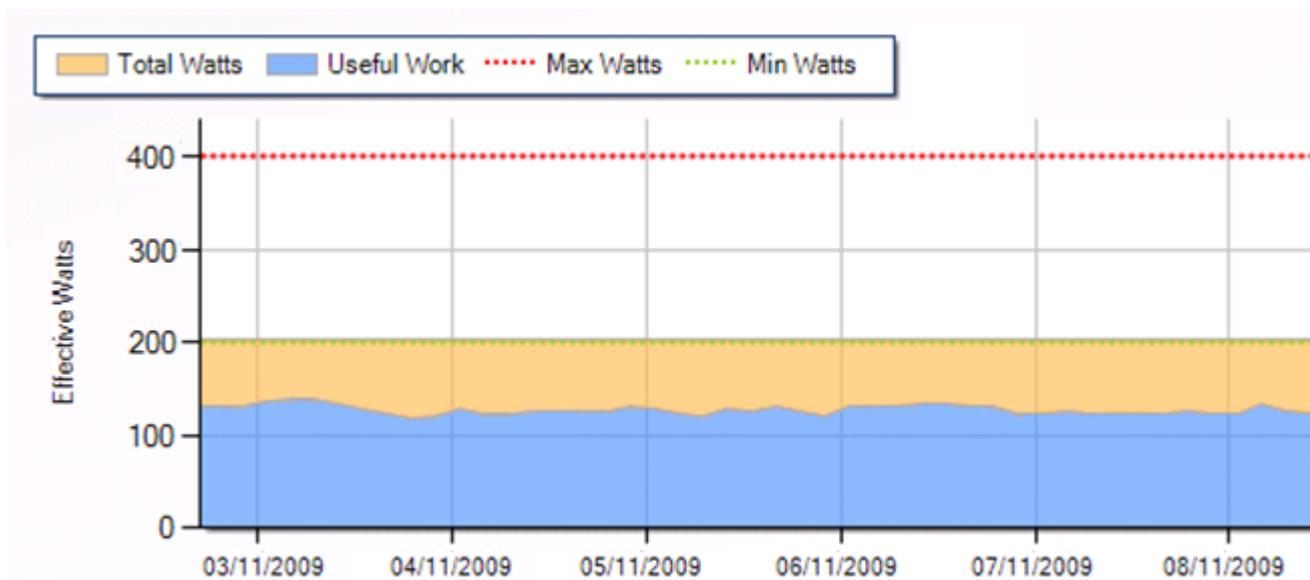
Useful work

But what constitutes “useful work”? We believe that if the server is busy doing the task for which it was bought and provisioned, then it is performing useful work, if it is busy doing anything else, however important, then it is performing non-productive work.

For example, a SQL Server doing SQL processing is performing useful work since users and/or applications will typically access SQL. The same server performing Anti-Virus scanning is not doing useful work since whilst AV scanning is important, it is not directly serving end users.

NightWatchman Server Edition uses cutting edge technology to analyze the actual workload of a server at the application level, and will show at a glance whether or not the server is performing useful work.

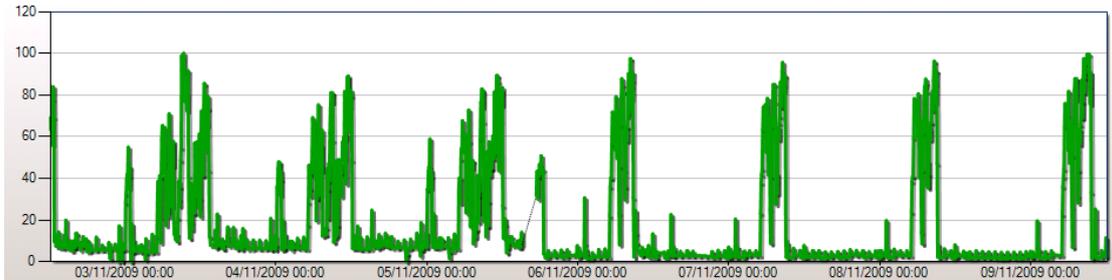
For example, below is an example of a well utilized server:



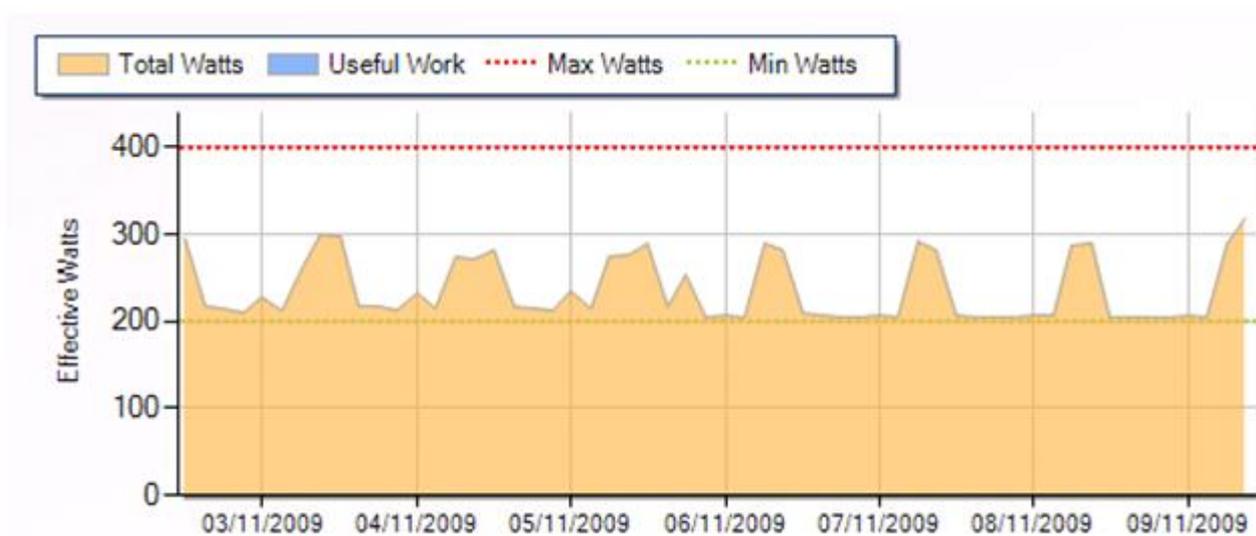
As you can see from the legend, the orange section represents the total Watts of power the server is consuming. This server is currently consuming approximately 200 Watts of power but at peak load could consume up to 400 Watts as denoted by the green and red dotted lines. The blue section represents the calculated useful work the server is performing. This server is considered well utilized as there is typically more than 50% useful work.

Now you see it..... Now you don't

From a typical systems monitoring solution, which only uses CPU utilization to show if a machine is busy or not, the server shown below appears fairly busy



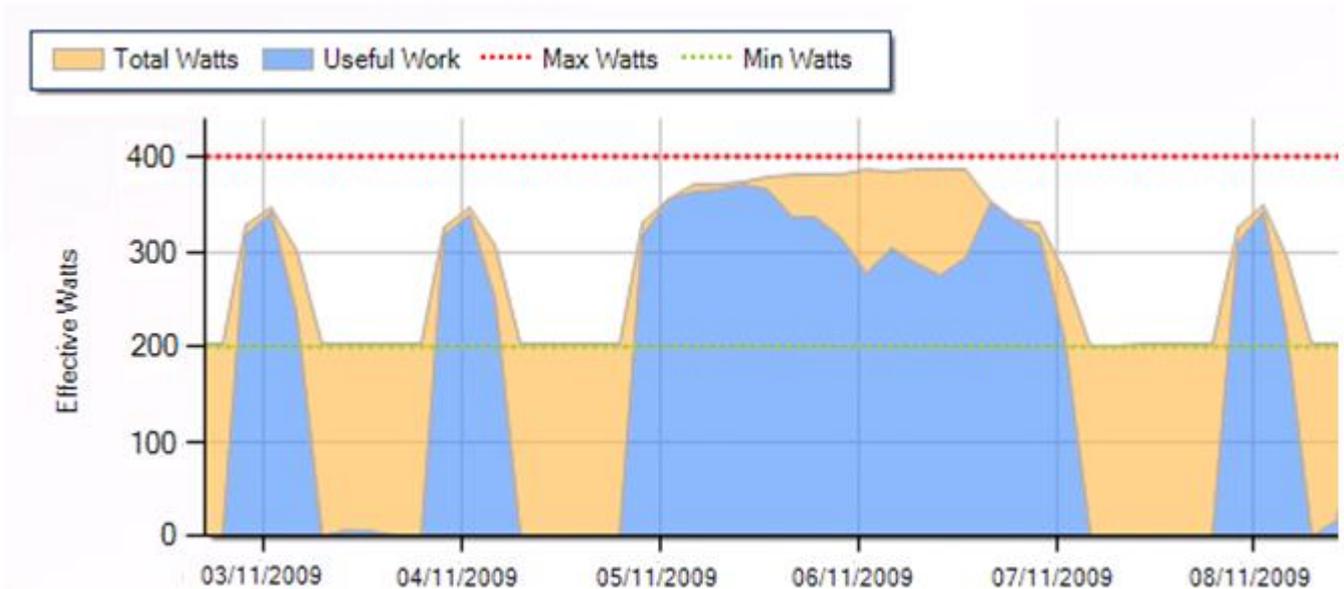
But when we look at the same server from the NightWatchman Server Edition console it tells a different story.



You can see that the entire graph is orange meaning that the server is only performing non-productive work. It is very likely that this server should be decommissioned, and NightWatchman Server Edition can produce analysis reports showing exactly which servers (both virtual and physical) are likely candidates for decommissioning.

Usefully useful

Finally, here's an example of a typical production server that experiences periods of both useful and non-productive work:



Since this server is actually performing useful work you wouldn't want to turn it off or decommission it, but NightWatchman server can still improve on the energy usage of the system by putting it into a 'Drowsy™' mode whenever it is not performing useful work, and we'll be explaining how that works in another document.

As you can see from the examples, useful work is presented in an easy to read format which really disguises the complex calculations which the product performs analyzing and collating the data necessary to generate these graphs.

For the first time, not only is it possible to identify how well a server is utilized but also whether or not a server is performing useful work or simply spinning its wheels with admin and maintenance tasks. We can finally tell when the lights are on but nobody is home.

All you non-productive servers, look busy....NightWatchman Server Edition is coming.