ZOMBIE/COMATOSE SERVERS REDUX

Jonathan Koomey* and Jon Taylor†

* Stanford University; †Anthesis

A report by Koomey Analytics and Anthesis

http://anthesisgroup.com/zombie-servers-redux/

April 3, 2017

ABSTRACT

In previous work, we showed that about thirty percent of the enterprise servers in our five facility, 4000-server sample were comatose, performing no useful computing over a six-month period in 2014. In this follow-up analysis, we assess the percentage of comatose (also known as zombie) servers in a sample taken in 2015 that covered about four times as many servers and twice as many facilities as were included in our original report, using a consistent methodology and data from TSO Logic.

Our analysis shows that about one quarter of physical servers were zombies in companies that had taken no action to remove them (which corresponds to the vast majority of companies running enterprise data centers). In addition, the data shows that about thirty percent of the virtual machines running on some of the physical servers (known as hypervisors) were also comatose, demonstrating that the same institutional and measurement problems that inhibit discovery and elimination of zombie physical servers also lead to significant numbers of zombie virtual machines.

Finding and eliminating comatose servers would save many enterprises money, but more importantly, taking that action would eliminate an unappreciated security risk. Zombie servers are unlikely to have the latest security patches, which makes them an open door to many enterprise data centers. If the monetary incentives are not enough to ensure prompt action, concern over cybersecurity really should.

INTRODUCTION

The problem of servers that use electricity but deliver no useful information services, (known as zombie or comatose servers) is one that continues to haunt the data center industry. Few companies can identify these orphaned servers, and many companies don't even know how many servers they have.

In previous work [1], we showed that about thirty percent of the enterprise servers were comatose in a 4000 server sample across five data center environments, performing no useful computing over a six month period in 2014. This finding mirrored earlier work by the Uptime Institute[1] and McKinsey and Company [2], which showed very low server utilization and significant percentages of zombie servers in enterprise data centers.

In this follow-up analysis, we assess the percentage of comatose servers in a sample four times larger than included in our original report (and covering twice as many facilities) using a consistent methodology and data from TSO Logic's data discovery tools. This report summarizes these new results, which includes an assessment of comatose physical servers as well as comatose virtual machines running on hypervisors.

UNDERSTANDING THE DATA

The data in the current samples were collected over six month periods in 2014 and 2015. TSO Logic anonymized the data carefully removing identifying information about each participating facility.

The data vary in many ways, and only by using precise terminology can we avoid confusion. **Table 1** gives definitions of some key terms.

The data associated with each sample will of course change over time. One key change is that the 2014 data don't split standard physical servers (which run only one operating system instance) from hypervisors (which run multiple virtual machines), while the 2015 data do.

The cohort comprising each sample may also change in composition over time: for example, the first sample from 2014 described five data center environments, but one of those facilities did not allow release of their data for 2015.

---

[1] https://uptimeinstitute.com/research-publications/asset/comatose-server-savings-calculator

Table 1: Key terms used in this study

Term	***Definition***
Standard server	Physical server that has 1 operating system and is not a hypervisor
Hypervisor	Physical server that supports multiple operating systems instances through virtualization (ie: Hyper-V or VMware)
Total servers	Aggregate of standard servers + hypervisors
Virtual machine (VM)	Number of operating instances on a hypervisor host
Active - Standard server	Standard server that showed sustained signs of activity (> 5% of the time*) over the report period including network, user, CPU, socket connection, or memory. Data points were taken every five minutes; data was consolidated into 15 minute internals and assigned an 'active or inactive' status. There were 5,760 of these 15 minute intervals during the six-month data capture period. When 288 or more of these 5,760 intervals (5% or greater) were designated as active the device as given an 'active' status for the 6-month period.
Active - Hypervisor	Hypervisor with VM's deployed and at least one of the VM's is active (> 5% of the time) over the report period
Active - VM	VM that showed sustained signs of activity (> 5% of the time) over the report period including network, user, CPU, socket connection, or memory usage
Idle - Standard server	Standard server that showed occasional signs of activity (compute and network) <5% of the time; Idle = Total – Active – Comatose
Idle - Hypervisor	Calculated value that represents the number of hypervisors that would be idle if the VM - host density was optimized across the aggregate of hypervisors
Idle - VM	VM that showed occasional signs of activity (compute and network) but for <5% of the time; Idle = Total – Active – Comatose
Comatose (Zombie)	Showed no signs of network, user, connection, memory or CPU activity in six months or more.
Comatose - Total servers	Standard servers + hypervisors that are comatose
Comatose - Hypervisor	Hypervisors that showed no signs of network, user, connection, memory or CPU activity in six months or more; OR no operating instances deployed
Comatose - VM	Number of virtual machines that have had no activity in the last 6 months
Comatose - Standard server	Standard server that showed no signs of network, user, connection, memory or CPU activity in six months or more

To make discussion of these changes more precise, we adopt terminology as follows:

1) We track each cohort by the year in which data were first collected for that cohort (e.g., 2014). If the composition of the cohort changes from one year to the next, we attach a prime, so the cohort for 2014 associated with data in 2015 (which dropped out one facility from the cohort) is listed as the 2014' cohort. If there is another change in that cohort (i.e., another facility dropping off in 2016) we would label that the 2014'' cohort.

2) We track the year in which the data are collected.

3) We track the type of data collected, such as the number of servers in each sample (N) or the percentage of comatose servers (P).

For each piece of data (D) we can write a term that looks like this:

$$D_{Year\ data\ collected}^{Year\ of\ cohort}$$

For the percentage of comatose servers for the 2014 cohort in 2014, that term would be written

$$P_{2014}^{2014}$$

For the 2015 data associated with the 2014 cohort, we lose one of the facilities in that cohort, so the term would be written

$$P_{2015}^{2014'}$$

Table 2 summarizes the relationship between cohorts and the year data are collected.

Table 2: Guide to sample terminology

	Year Sample Collected				
Cohort Name	**2014**	**2015**	*2016*	*2017*	*2018*
2014 (facilities 1 - 5)	**2014**	**2014'**	*2014'*		
2015 (facilities 6 - 12)		**2015**	*2015*	*2015*	
2016 (facilities 13 - n)			*2016*	*2016*	*2016*

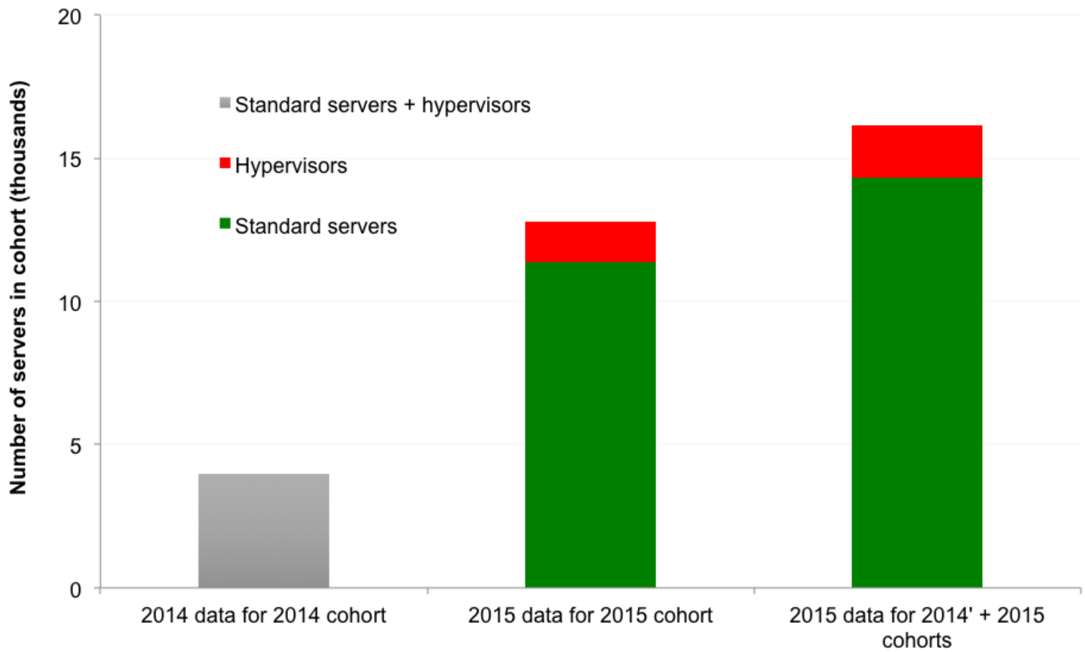
Note: Italics indicates future samples. Prime (') indicates that cohort changed from previous year, so 2014' indicates that the 2014 sample lost one facility in 2015.

Figure 1 compares the 2014 server sample size, (including facilities 1 to 5) with the number of servers in the new facilities in the 2015 sample (facilities 6 to 12). The third bar sums the sample sizes of the 2014' cohort (less one facility from 2014) and the 2015 cohort.

These can be described in the following terms (which match up in sequence left to right with the bars in Figure 1):

$$N_{2014}^{2014}\ , N_{2015}^{2015}\ , \ N_{2015}^{2014'} + \ N_{2015}^{2015}$$

Figure 1: Numbers of physical servers in the 2014 and 2015 cohorts

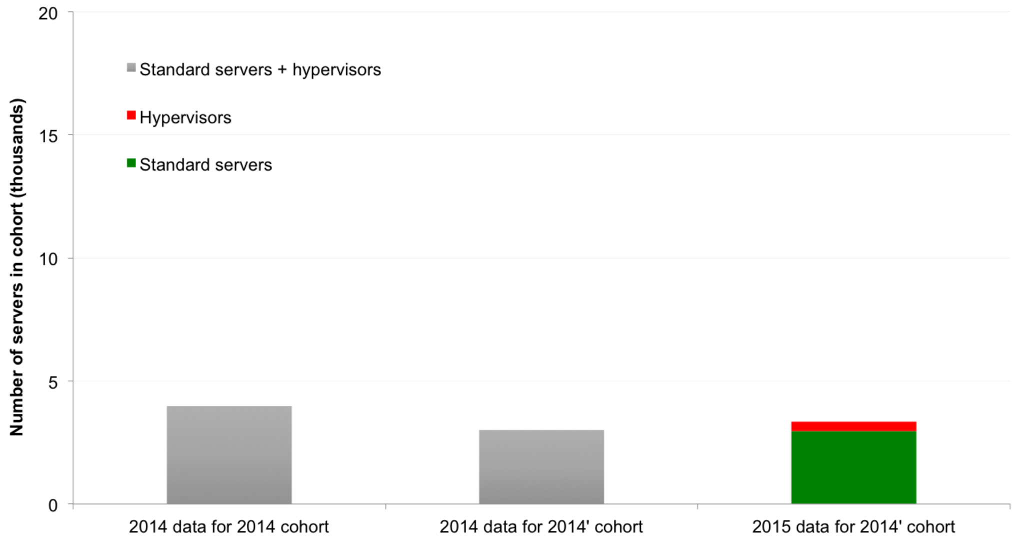


Key: 2014 data for 2014 cohort = N_{2014}^{2014}; 2015 data for 2015 cohort = N_{2015}^{2015}; 2015 data for 2014' + 2015 cohorts = $N_{2015}^{2014'} + N_{2015}^{2015}$

Recall that the 2014 data don't split the standard servers from the hypervisors, as shown by the single bar on the left hand side in Figure 1. The combined sample in 2015 is about four times bigger than the 2014 sample.

Figure 2 shows the effect on sample size of omitting the one facility that dropped from the sample, which reduces the 2014 server sample size by about 1000 (shown in the drop from the leftmost bar to the middle bar).

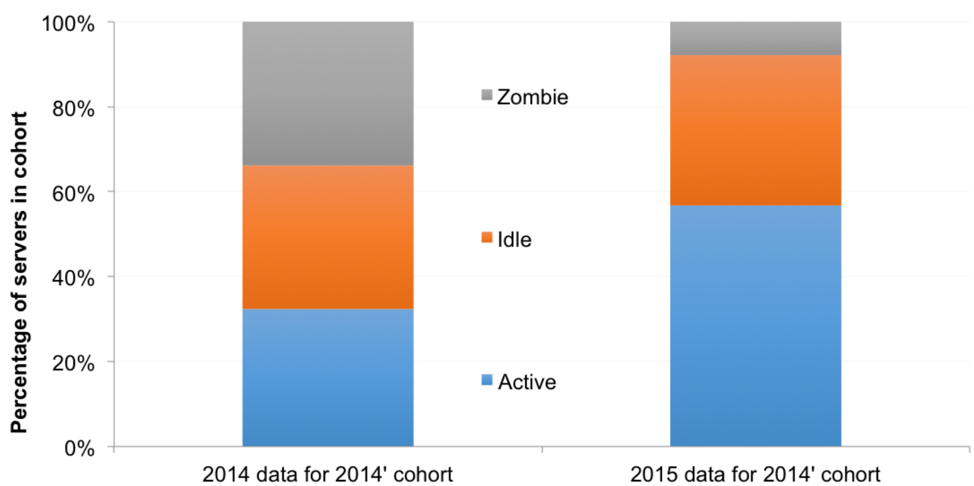
Figure 2: Numbers of physical servers adjusted to reflect missing data from 2014 sample in 2015 (for the 2014' cohort)



Some of the remaining 2014' facilities added servers from 2014 to 2015, which is reflected in the increase in the rightmost bar (2015 data for 2014' cohort) in Figure 2 compared to the middle bar.

Figure 3 compares the percentages of active, idle, and comatose servers in the 2014' cohort in 2014 and 2015 (see Table 1 above for definitions). Zombie servers drop from 34% of the sample in the 2014' cohort in 2014 to 8% in 2015, reflecting what happens when management realizes just how much money is being wasted on zombie servers and takes appropriate action to eliminate that waste.

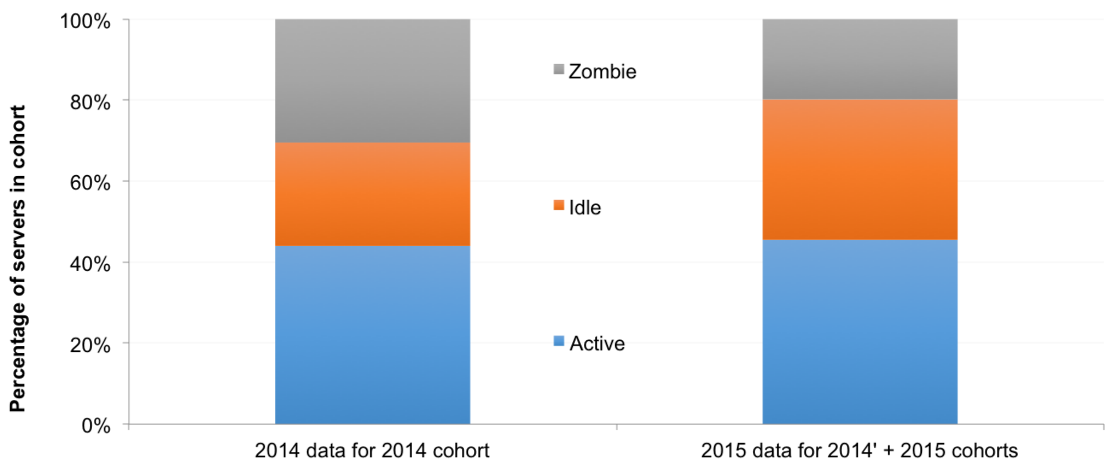
Figure 3: Percentages of physical servers by usage level for the 2014' cohort in 2014 and 2015



The drop in the percentage of zombie servers between the 2014’ cohort in 2014 and 2015 shows that companies took actions to eliminate the zombie server problem when presented with clear evidence of that problem’s magnitude from TSO Logic’s discovery tools.

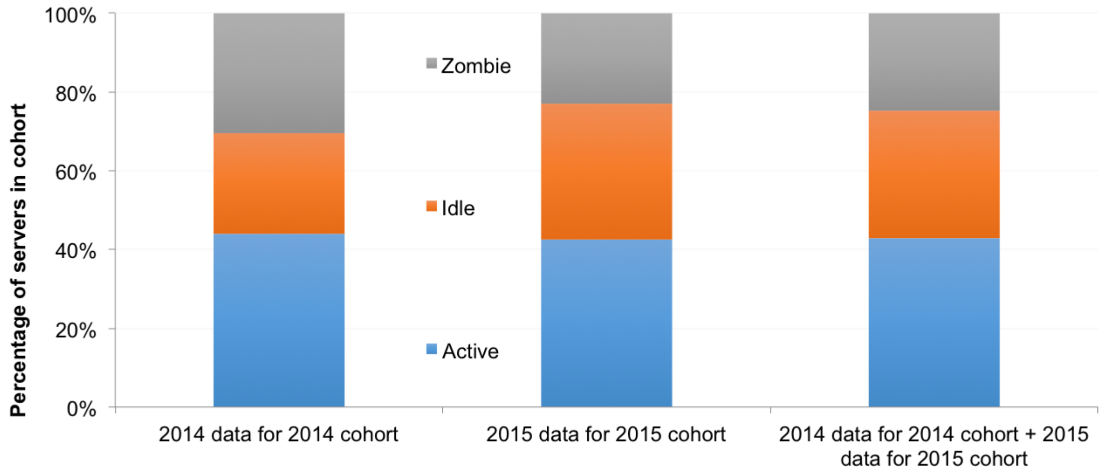
Figure 4 plots the percentage by usage category in each cohort, contrasting the 2014 data for the 2014 sample with the 2015 data for the sum of 2014’ and 2015 samples. The big shift in the percentage of comatose servers from the 2014 cohort in 2014 to the 2014’ cohort in 2015 results in a lower estimate of the percentage of comatose servers in the combined sample in 2015 (about 20%) than we might expect if we were measuring usage behavior in companies who had not taken action against comatose servers. One of our goals is to assess the percentage of comatose servers for typical enterprises, who for the most part have taken little or no action against comatose servers. For this reasons we need to adjust the data a bit.

Figure 4: Percentages of physical servers by usage level for 2014 data for the 2014 cohort and 2015 data for the sum of 2014’ and 2015 cohorts



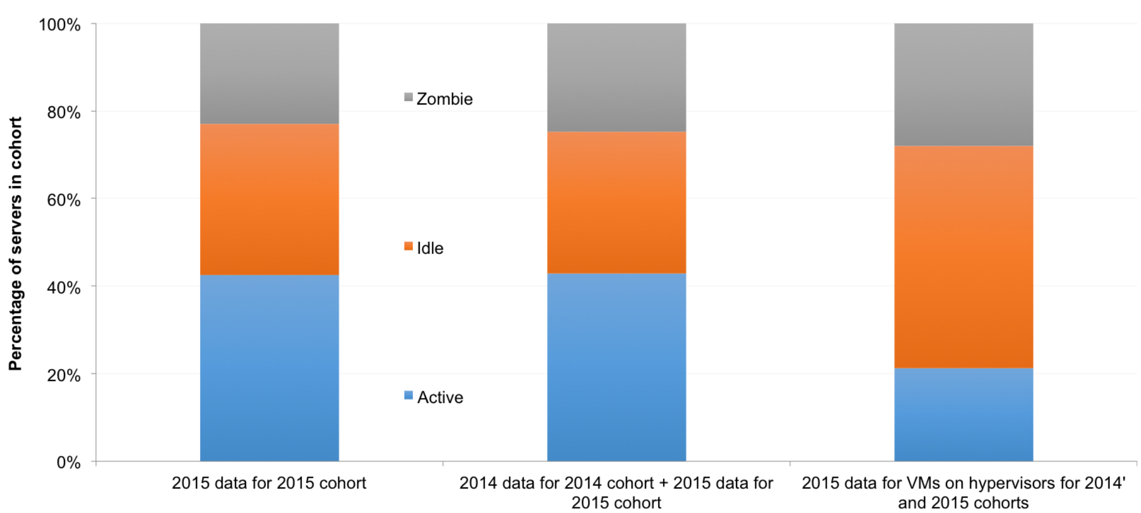
If we treat the 2014 data for the 2014 cohort and the 2015 data for the 2015 cohort as one sample, we would in essence be assuming that the corrective actions taken by the 2014 cohort did not occur, and it would give us a clearer picture of usage patterns in companies who haven’t yet attacked the zombie server problem. That assumption gives us the graph shown in **Figure 5**, which compares the 2014 data for the 2014 cohort, the 2015 data for the 2015 cohort, and the data for the newly combined sample that sums data for the 2014 and 2015 cohorts. In the 2014 sample, zombie servers comprise about 30% of the total, while for the 2015 sample comatose servers represent 23% of the total. In the corrected sample, which combines the 2014 data and 2015 data, zombie servers are 25% of the total in 2015.

Figure 5: Percentages of physical servers by usage level for 2014 and 2015 samples individually, then combined as one sample



These data confirm the existence of significant numbers of zombie physical servers, but they also reveal that the same institutional and measurement problems that inhibit discovery and elimination of zombie physical servers also lead to significant numbers of zombie virtual machines. **Figure 6** compares the percentages of physical servers by usage category for the 2015 cohort in 2015 from Figure 5, the combined 2014 and 2015 sample from Figure 5, and the percentage of virtual servers (VMs) by usage category in the 2014' plus 2015 sample (in the rightmost bar).

Figure 6: Percentages of physical and virtual servers by usage category for different cohorts in 2015



Many industry experts believe that virtualization is a silver bullet for achieving high server utilization (and better use of data center capital) but the results of Figure 6 contradict this belief. About 30% of the virtual machines running on the hypervisors in our 2015 sample were comatose.

Thus, almost one third of the capital allocated to those machines is generating zero monetary return, which is surprising in part because hypervisors tend to be much more expensive than standard servers on a per server basis. Interestingly, about 10% of the hypervisors in the combined 2014' plus 2015 samples for 2015 were zombies, indicating problems with the management of both the VMs and the physical assets on which those VMs are housed. Virtualization without improved measurement technology and altered management and institutional changes is not a panacea[2]!

Aside from wasting money, zombie servers are clearly a security risk. A server that isn't doing anything useful is also not likely to have the latest security patches, which makes zombie servers an open door into many enterprise data centers. For this reason alone identifying these servers should be a priority.

CONCLUSIONS

Until much larger sample sizes become available we must be cautious in drawing precise conclusions. In broad terms, this study confirms that zombie servers are an important issue that is still underappreciated by the data center community.

The percentage range of zombie servers in the combined data set (2014 + 2015) is lower in this study (about 25% instead of 30% in the previous study). Interestingly, the virtual machines in our 2015 data set show that a large percentage are zombied (about 30%), indicating that management failures affect physical and virtual servers alike.

Our data indicate that one quarter to one third of data center investments are tied up with zombie servers and virtual machines, and that capital (which no doubt totals in the tens of billions of dollars in the US alone) is generating zero financial return. Chief financial officers everywhere should leap to fix this problem, because the potential for improving operations and management of these facilities is vast [3, 4, 5, 6, 7] and because zombie servers represent a security risk.

The tools to identify zombie servers are only now becoming widely available and the institutional problems that lead to the existence of many zombie servers also inhibit their discovery. Without visibility into the scale of these wasted resources and significant institutional changes, this waste of data center capital will continue to bedevil the industry.

ACKNOWLEDGMENTS

[2] Another issue that is outside of the scope of our study is that virtualization in the absence of automated deployment of VMs is much less beneficial than virtualization with automation (which is common in cloud deployments). Automated VM deployment allows for much more rapid deployments than in the past, and that speed to market is a key benefit of automated virtualized cloud environments.

We would like to thank Aaron Rallo at TSO Logic for sharing anonymized data with us for this project, and for both Aaron and Andrea Majorski of TSO Logic for contributing comments and insights for the analysis. Nicole Peill-Moelter of Akamai also contributed helpful comments. This project was conducted by Jonathan Koomey and Jon Taylor on their personal time, with no external funding source. All errors and omissions are the responsibility of the authors alone.

REFERENCES

1. Koomey, Jonathan, and Jon Taylor. 2015. *New data supports finding that 30 percent of servers are 'Comatose', indicating that nearly a third of capital in enterprise data centers is wasted.* Oakland, CA: Anthesis Group. [http://anthesisgroup.com/wp-content/uploads/2015/06/Case-Study_DataSupports30PercentComatoseEstimate-FINAL_06032015.pdf]

2. Kaplan, James M., William Forrest, and Noah Kindler. 2008. *Revolutionizing Data Center Efficiency* McKinsey and Company.

3. Shehabi, Arman, Sarah Smith, Dale A. Sartor, Richard E. Brown, Magnus Herrlin, Jonathan G. Koomey, Eric R. Masanet, Nathaniel Horner, Inês Lima Azevedo, and William Lintner. 2016. *United States Data Center Energy Usage Report.* Berkeley, CA: Lawrence Berkeley National Laboratory. LBNL-1005775. June 27. [http://eta.lbl.gov/publications/united-states-data-center-energy-usag]

4. Koomey, Jonathan. 2014. *Bringing Enterprise Computing into the 21st Century: A Management and Sustainability Challenge*, March 17, 2014 [http://www.corporateecoforum.com/bringing-enterprise-computing-21st-century-management-sustainability-challenge/]

5. Koomey, Jonathan. 2016. "Applying the Scientific Method in Data Center Management." In *Data Center Knowledge*. March 9. pp. [http://www.datacenterknowledge.com/archives/2016/03/09/applying-scientific-method-data-center-management/]

6. Koomey, Jonathan. 2016. "Three Pillars of Modern Data Center Operations." In *Data Center Knowledge*. February 2. pp. [http://www.datacenterknowledge.com/archives/2016/02/02/three-pillars-modern-data-center-operations/]

7. Koomey, Jonathan, and Patrick Flynn. 2014. "How to run data center operations like a well oiled machine." In *DCD Focus*. September/October 2014. pp. 81. [http://goo.gl/7sJHZb]