Realtime publishers

The Essentials Series

Why You Need to Defragment

sponsored by

by Greg Shields

Copyright Statement

Article 1: Fragmentation Is a Problem!

Why do we defragment? Simply put, because we must!

Data fragmentation on a computer's disk drives quickly creates a major source of performance loss. It increases the time required to accomplish every task on your system, including launching applications, working with data, interacting with page files and hibernation files, all the way to the otherwise-innocuous startup and shutdown activities. It adds an unnecessary layer of complexity to the storage of files and folders, shattering the contiguous storage of on-disk data into dozens or even hundreds of individual pieces. Its constant reordering makes data less reliable to restore in the case of a loss and more difficult to reassemble when needed for processing.

Fragmentation on the disks of Windows servers and workstations has been around since the very first disk. It is a necessary evil of disk-based storage, and is an almost unavoidable consequence of the ever-present process of reading, writing, deleting, and writing again to a computer's storage. Left unmanaged, virtually every time a piece of data is touched by a Windows computer's file system, its action forces the creation of yet another fragment.

In essence, if you've worked with the Microsoft Windows operating system (OS) for any period of time, you've heard of this problem. But in hearing about fragmentation, do you truly understand its meaning? Do you recognize why fragmentation is an endemic problem on each and every Windows computer, one that must be continuously managed if it is to be kept under control? Were you aware of the true scope of fragmentation, and how many fragments an average knowledge worker's desktop produces each and every week? If not, read on.

Fragmentation, the Silent Killer

Testing has shown that an average desktop, one commonly used in a business network environment, can accumulate upwards of 12,000 individual fragments per week (Source: http://downloads.diskeeper.com/pdf/Real-Time-Defrag-Whitepaper.pdf). This number is cumulative, meaning that additional weeks add additional fragments over the top. The net result is a linearly-scaling level of fragmentation on a computer's hard drive that must be managed. Without tools to reassemble fragments into contiguous files on disk or prevent their occurrence in the first place, this problem will eventually scale to slow the overall performance of that system.

Fragmentation is a naturally-occurring phenomenon that is associated with the storage of file system data on a computer. The process of fragmenting a file is not something that can be stopped in a file system without the assistance of specific third-party algorithms. To combat its effects, a separate process must be incorporated to manage the reassembly of file fragments in parallel with a file system's operation.

Data fragmentation occurs when a unit of data on a computer's hard disk is broken up into many pieces. This happens due to the natural use and expansion of data within a computer system. Computer disks store data linearly, meaning that a unit of data is laid down in a contiguous fashion by a disk's head. The rotation of that disk causes the head to pass by the disk's platter, reading and writing data across that disk's sectors and tracks. This is represented in Figure 1, where disks at three points in time are shown as rectangles. In the top representation, File A is written to the disk. In the next unit of time, File B is written to the disk as shown in the middle rectangle.

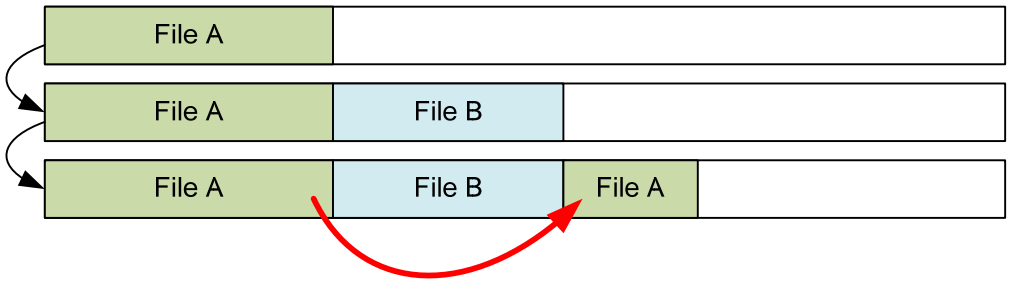

Figure 1: When File A must expand in size, it must fragment to the next available area of storage space.

At this point in the example, the two files remain contiguous on disk because they were initially created and have not yet experienced growth in size. That growth is represented in the bottom rectangle as the third period of time. Perhaps File A was a Microsoft Word document that needed a bit of extra work. Maybe File A was a system DLL that was updated by a patch or a system routine. In either of these cases, this additional processing of File A required an additional bit of space on disk; however, no contiguous space is available. Thus, File A must be fragmented to the next available piece of space, which is located after File B.

This exact situation is what happens upwards of 12,000 times per week on each and every hard drive in your computer. The daily operations of a computer system require the constant expansion of files, the deletion of files, and the placement of files into open spaces that are made available. As this process iterates, individual files can become fragmented dozens or hundreds of times.

The Cost of Fragmentation

The result is that a single file can require multiple disk passes to be completely read into memory for processing. Rather than reading an entire file at once, the disk's head must locate and read each individual fragment, while at the same time reassembling each of these fragments into useable data. As the level of fragmentation increases, the processing overhead associated with these actions dramatically impacts your computers' performance.

How much performance is lost through this accumulated process? Studies show that once a disk is defragmented, the entire system can see a performance gain of up to 80%, with an average realized benefit of 10% to 20% (Source: http://downloads.diskeeper.com/pdf/The-Impact-Of-Disk-Fragmentation-On-Servers.pdf). Obviously, the improvement in performance is directly related to the amount of fragmentation that can be eliminated, with more fragmentation causing more slowness problems.

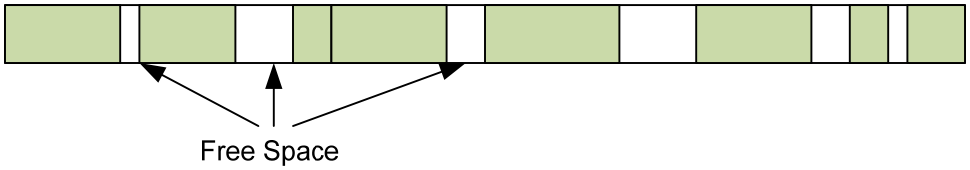

Figure 2: Accumulated fragmentation also impacts the availability of free space on a disk.

A secondary set of problems that grows worse as the level of fragmentation increases has to do with your systems' available free space. Figure 2 shows an example of a disk that has been naturally fragmented due to the typical operations of a Windows OS. There, you can see how the iterative writing, expansion, deletion, and re-writing of files has forced the file system to create "holes" of available disk space. Over time, the count of these holes grows while the size of each hole actually goes down. This reduction in size of free space segments impacts the performance of future writes, because any future writing of files automatically starts in a fragmented state. In effect, *fragmentation begets more fragmentation*.

Solving the Problem

The net result of these factors means that unmanaged fragmentation directly impacts the ability for your users to get their jobs done. As a natural process of the Windows OS, fragmentation isn't going away. And without the right defragmentation tools in place, your users will experience unnecessary slowdowns in performance, your servers will service their clients with reduced effectiveness, and you may find yourself purchasing new and faster hardware that needn't be a part of your budget.

The next two articles in this series will discuss just those problems. Article two will further hone in on the fact that *You Need to Defragment!*, explaining where and why fragmentation impacts system performance and how good practices in defragmentation improve your overall network infrastructure. Article three continues the conversation by answering the question *Doesn't Windows Have This?*, explaining why native OS tools are insufficient to truly get the job done.

Article 2: You Need to Defragment!

Can we all agree that there's nothing more frustrating than a slow computer? You've probably experienced the following situations, because they can all be common to your daily interaction with the Windows operating system (OS):

- You need to finish that spreadsheet before heading home to dinner and family, but instead you're watching the hourglass tick by.
- Maybe it's a quick email check before boarding that flight, but you forego the opportunity because your laptop takes 4 excruciating minutes to boot.
- Or, you're stacked with meetings and PowerPoint charts but find yourself in a waiting pattern as you reboot that un-cooperating conference room PC.

In these and dozens of other situations, you're at the mercy of your computer's processing. When it doesn't perform to the needs of your daily workflow, it can feel like you're working for it instead of it working for you. In virtually all these scenarios, that computer's lack of performance can be directly impacted by its level of fragmentation.

Fragment-Less Is the Goal

Article one in this series outlined the problem of fragmentation. It explained how fragmentation is a naturally-occurring side effect of a computer's disk activity. As something that cannot be naturally stopped, disk fragmentation must instead be managed to keep its spread from slowing your processing.

To that end, there are a number of established best practices associated with managing defragmentation as well as tools that enable proactive defragmentation. Although not all solutions are created equal, smart organizations select those with the right set of capabilities which ensure fragment-less systems both in the desktop and the data center. One long-held mechanism to accomplish this relates to the window in which defragmentation can occur.

Continuous > Scheduled

Traditional defragmentation solutions offer options for scheduling the defragmentation "pass" on your systems. This pass needs to be scheduled to occur at off-hour intervals, as its impact on system resources can be dramatic. The reassembly of file and folder fragments tends to be of great impact to the file system as well as the disk subsystem as a whole. Its processing can require a substantial amount of processor and memory resources as the defragmentation pass completes. These resources are necessary due to the multi-step process associated with analyzing a disk drive, looking for files, and correctly assembling them into a logical order. Should these activities occur without proper resource throttling and poor scheduling, the defragmentation process itself can be a major impact on your users' experience.

Yet although this concept of scheduled defragmentation has been a de facto norm for many years, many defragmentation solutions today leverage an alternative approach to optimizing file structures. Eliminating the schedule altogether, these solutions instead opt for a continuous approach to finding and fixing fragments.

Consider how this alternative approach improves the entire process. Article one discussed how the sheer number of fragments grows dramatically as a computer system is used. Computers with larger numbers of writes and a greater count of files tend to have a larger quantity of fragments. Thus, once the time interval goes by between fragmentation passes, the defragmenter starts at a disadvantage: To return a volume to a defragmented state, it must "catch up to" and eventually "get ahead of" the data processing of the system.

This problem tends not to be as challenging with desktop systems. Users of these systems often don't use them 24 hours a day. Thus, a natural period exists when processing is low and defragmentation can catch up. However, scheduled jobs on desktops can be problematic when users don't leave those systems powered on during the scheduled interval. Depending on the solution available, a powered off system can either miss the defragmentation schedule or be forced to run it shortly after the system is powered back on—and the user is ready to make use of it again.

With servers, the problems surrounding this approach grow even more insidious. Imagine the typical file server or database server, which tends to process its workload during the business day. High resource use actions for servers—such as patching, backups, and defragmentation—tend to collect during the evening hours. The co-processing of these intensive tasks over the short period away from business hours can aggregate to dramatically increase the overall time to accomplish each.

Contrast this situation with the continuous approach. Here, a computer's file system is always being monitored by the defragmentation solution. When fragments appear, those fragments are handled almost immediately. Today's enterprise defragmentation solutions leverage the interstices between user requests to accomplish the defragmentation process. As a low-level service that occurs in combination with the file system's writes, this incremental approach ensures that your disks remain defragmented and highly optimized at all times.

Proactive > Continuous

Yet even this continuous approach remains a reactive band-aid to a never-ending problem. Defragmentation products that rely exclusively on even a continuous approach find themselves working to resolve a problem that could be best solved by ensuring it never happens in the first place. This modern "proactive" approach to defragmentation dramatically changes the ways in which fragments are managed by a computer system.

For example, consider the situation where a new file is added to a perfectly defragmented disk. Even though this disk is completely free of fragments, "holes" of free space tend to lie across multiple areas on the disk. When this new file is added, the computer's file system attempts to locate a hole of free space within which to store the file. Using native tools alone, that file system is likely to store the file into a hole that isn't quite large enough to store the entire contents of the file. Immediately, a fragment is created as the file's contents are spread across multiple holes.

Using the continuous approach, once the file system completes its write, it is the job of the defragmentation engine to locate and reposition that file (as well as others that surround it when necessary) into a location where it is no longer fragmented. Using this process, the defragmentation engine is constantly forced to react to poor decisions that are made by the file system. When that file later expands, this doubling of effort repeats itself all over again.

Contrast this situation with one where the defragmentation engine and the file system work together instead of at odds with each other. Using this approach, any new file can be automatically written to the system's disk in such a manner so that little or no fragmentation occurs. File writes and expansions are compensated for by the defragmentation engine with the support of the file system itself. In essence, when using the proactive approach, a computer's disk largely prevents file fragmentation at any point. Solutions that leverage the proactive approach accomplish the same goal of a fragment-less system but with far less effort, impact on system resources, and the assurance that most file writes can be done without fragmentation ever occurring.

Fragmentation Impacts Everything

Ultimately, the sole purpose of defragmenting a computer is to increase performance. That point has been repeated thoroughly in this series already. But what kinds of processes are impacted by fragmentation? What types of user activities can be improved by the implementation of effective enterprise defragmentation? The first set of areas worth reviewing relates to the individual desktops and laptops of your users themselves. Consider the following user activities that are improved through the assurance of an always-optimized file system:

- *Slow application and OS response time.* Testing using the PCMark performance benchmarking tool has shown that a fragmented file system can have a dramatic impact on desktop performance (Source: http://downloads.diskeeper.com/pdf/NSTL_20Tests_20Diskeeper_20vs_20Built_20In.pdf). The running of this tool generates a metric that aggregates overall system performance and is intended to be used in comparison with other numbers from the same tool. Here, fragmented desktops scored a 4763.2, while those which leverage the services of external defragmentation solutions scored a 5484.6. Thus, the net gain in overall system performance in this single test was around 14%.

- *Increased time to power on.* It can be argued that one of the most resource-intensive activities on any desktop or laptop system is related to its powering on. The bootstrapping as well as shell and user interface-instantiation processes require the involvement of numerous system components, all of which must occur in a very short period of time. Similar testing using Microsoft's Xperf.exe tool has shown that a fully-optimized disk drive can improve power on performance by an average of 3 to 5 seconds. Although this may not be dramatic for desktop users, this time savings is a boon for laptop users. This improvement in performance also extends to the hibernation process, whereby a laptop is put to sleep and later revived without requiring a full power-on process. As this process requires the creation and maintenance of a large hibernation file, its fragmentation further increases the process to revive a sleeping laptop.
- *System crashes and freezes.* As discussed in the first article, the process of fragmentation quickly spreads individual pieces of data into multiple locations. This widespread shattering of individual data files increases the chance that their later reassembly may fail, or may force a system freeze during the reassembly process. Eliminating fragmentation on a file system removes this variable from file systems, ensuring that files can be gathered from disk in a contiguous fashion.
- *Performance impact to existing enterprise services.* Lastly, the impact of fragmentation has a dramatic effect on other enterprise services, notably those that have a high reliance on disk and file system resources. Consider common business services such as antivirus and anti-malware. The mission of these agent-based solutions is to monitor the file system and processing for the potential intrusion of malicious code onto the system. Both real time and scheduled scans are often required for full assurance, so their processing is directly affected by the performance loss associated with data fragmentation.

The impacts on individual desktops and laptops are important to ensuring high levels of IT customer satisfaction. Yet the role of defragmentation doesn't stop at the data center's doorway. Inside that data center are another set of Windows OSs that operate in a server role. They too are impacted by the performance loss associated with file fragmentation, although any performance reductions here are experienced by a much larger audience than with any single desktop or laptop. Consider their additional situations:

- *Decrease in overall performance, particularly with very large files.* Implementing a proactive approach ensures that files make their way to disk in a non-fragmented state, and there is little to no need for later defragmentation to occur. Reactive defragmentation can be affective to resources on servers with very large storage requirements. It is particularly resource intensive when files are exceptionally large, such as those used by virtual machines or databases. Leveraging a defragmentation solution that uses the proactive approach means eliminating this performance impact to your servers.

- *Reduction in backup performance and increase in backup windows.* Files and folders must be reassembled before they can make their way to tape. Thus, the incremental process of archiving copies of your servers' data can take dramatically longer when not properly optimized for performance. This delay is further problematic as it increases the window of time required to complete backups, potentially complicating other off-hour tasks required in the data center.
- *Reduced ability to undelete files.* When a file is fragmented into multiple pieces, that file is spread across the disk's area. In cases where files are accidentally or maliciously deleted and require un-deletion, such a fragmented file has a dramatically lower chance of a successful restore. This happens because its individual pieces have a much greater likelihood of being overwritten by other data after the deletion event. This chance grows as the amount of time between the deletion event passes, giving the file system more opportunities to overwrite pieces of the original file.
- *Dramatically reduced performance of virtual machines.* With their entire disk subsystem consolidated into single files on another server's disk, the processing of virtual machines is exceptionally dependant on file system performance. When the very large disk files associated with virtual machines grow fragmented—a situation that is particularly problematic when virtual machine disk files are configured to grow as needed—the resulting reduction in the virtual machine's performance can be dramatic. This is the case for both the virtual machine's file on its host disk as well as fragmentation within the virtual machine's disk drives.

Defragmentation Equals Performance

As you can see through the examples discussed in this article, defragmentation is indeed primarily about your systems' performance. By implementing a policy of defragmentation that corresponds to established best practices and modern approaches, you will ensure the highest levels of performance for the systems in your network. This makes users happy while reducing the need for costly and unnecessary hardware upgrades.

Yet, throughout all this discussion, the question is begged *Doesn't Windows Have This?* A not-inappropriate question, the Windows OS does indeed arrive with its own built-in defragmentation solution. The third and final article in this series will discuss compelling characteristics of that native solution in relation to the capabilities your business needs.

Article 3: Doesn't Windows Have This?

Of course it does. But as with many other things in life, with Windows' onboard defragmentation engine, *you get what you pay for.* Microsoft's built-in disk defragmentation tool is a solution that was originally obtained from its third-party ecosystem. Starting in the early 1990s, Microsoft ported this third-party code into its operating system (OS) as a built-in solution for accomplishing basic defragmentation operations.

However, the codebase incorporated with this port remains dramatically different than those available through today's third-party software vendors. Although the core performance of this solution is visibly improved in newer OSs such as Windows Vista and Windows 7, its implementation on Windows XP systems simply does not provide the level of defragmentation required by most enterprises.

Even with Windows Vista, Windows 7, and Windows Server 2008 R2, Microsoft's defragmentation implementation today remains only a stopgap measure to prevent the grossest levels of fragmentation. As an example of this, reference Figure 1, where two representations of a Windows file system are presented.

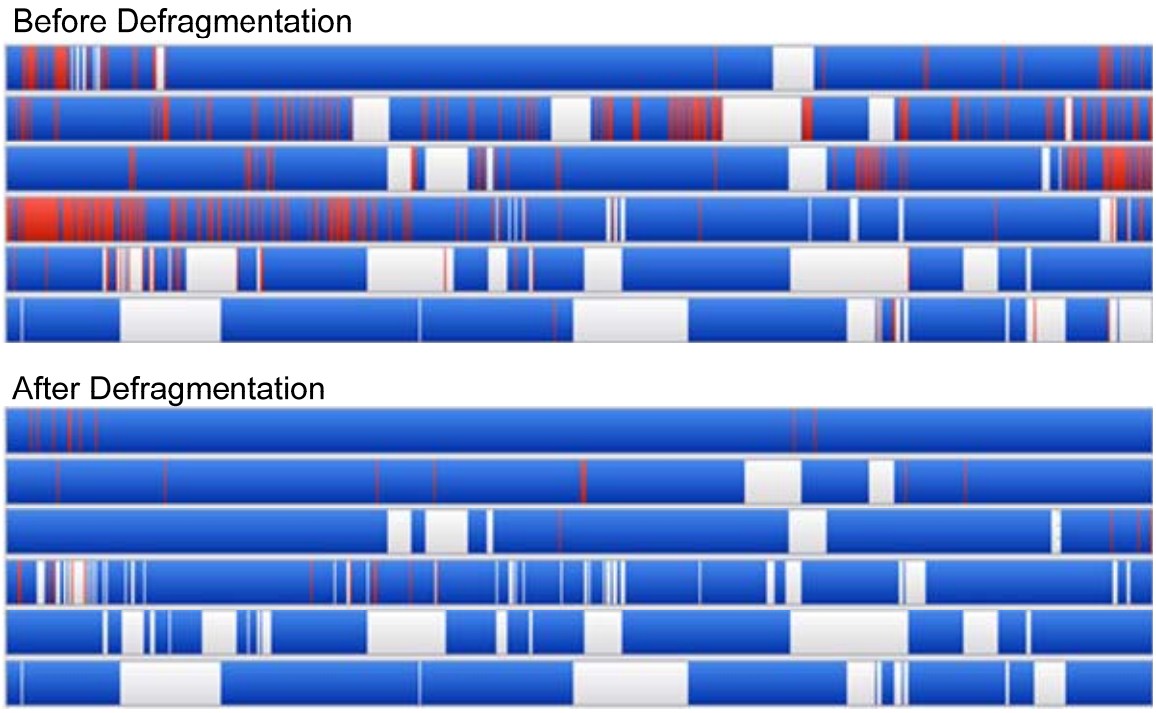

Figure 1: File performance with the standard Vista defragger (top) and after using a third-party defragmentation solution.

In this figure, the image on the top represents the level of fragmentation on a Windows Vista computer that has used only the native defragmentation solution. This computer has been in operation for nearly 2 years, using only the native weekly defragmentation available in Windows Vista.

You'll notice here that a number of areas are marked in red. These correspond to files and folders that have not been fully defragmented and are not operating at full efficiency. Even though the native defragmentation solution was scheduled to occur on a weekly basis, that defragmentation pass was unable to fully complete its mission. Compare this graphic with the alternative on the bottom, which was taken immediately after completing a defragmentation pass on this same computer using a third-party defragmentation solution. Here, you'll see that the number of non-optimized files is dramatically reduced through just a single pass of the third-party solution.

Limitations of the Native Defragger

Article two of this series discussed how today's conventional wisdom associated with defragmentation has dictated a proactive approach. Using the proactive approach, the level of resources required by the defragmentation engine is dramatically reduced. Defragmentation simply isn't allowed to exist on the system, which means that proactive management also ensures a fully-optimized file system.

In contrast, the native Windows defragmentation solution leverages a less-effective scheduled approach to its processing. By default, it invokes a defragmentation pass every Wednesday at 1:00am on desktops, which can directly impact system performance while it goes through its machinations. Due to the architecture of the Windows scheduling engine, if this scheduled pass is missed due to the machine being powered off, the pass will instead occur at the next power on.

Further, the native Windows defragging solution is limited to online operations only. There are some files in the Windows file system that cannot be optimized while the system is powered on. These files, such as the system paging file and hibernation file can accumulate their own levels of fragmentation over time, especially when configured for growth. One result of this limitation is an inability to consolidate free space across the computer's hard disk, leaving the aforementioned "holes" of free space on a defragmented disk. Alternative solutions that enable proactive and continuous defragmentation are necessary for these files to be fully optimized.

Impacts on Servers

It is not a well-known fact that Windows' native defragmentation solution is disabled by default on Windows Server 2008. But before you go about enabling it on all your systems, consider the impact: Enabling that schedule can have a dramatic impact on performance during its initial and even future passes. This fact means that many business networks are likely operating their servers with exceptionally high levels of fragmentation, potentially causing a major impact on their server operations.

You cannot simply enable this schedule without expecting some ramifications. Although the native Windows solution incorporates limited process throttling to prevent resource overuse, that throttling is reactive in nature. As such, to protect yourself against a measure of pain, consider the use of third-party solutions that leverage proactive solutions for resource overuse prevention before ever turning on Microsoft's native solution on your servers.

Nowhere is this more dramatic than on servers with very large volumes. These volumes, which may measure in the hundreds of gigabytes or even terabytes, have special needs due to the sheer size of their data storage. As the defragmentation process requires involvement from processing and memory resources in order to accomplish its optimization, servers with very large volumes should also consider the use of external solutions that are designed to scale.

Impacts on Management

Finally, there are two useful management elements that are missing from the native defragmentation solution in the Windows OS. The first of these is a user interface (UI) that provides the right level of detail to users. As you can see in Figure 2, the Disk Defragmenter wizard in Windows 7 is very limited in the information it presents to its users.

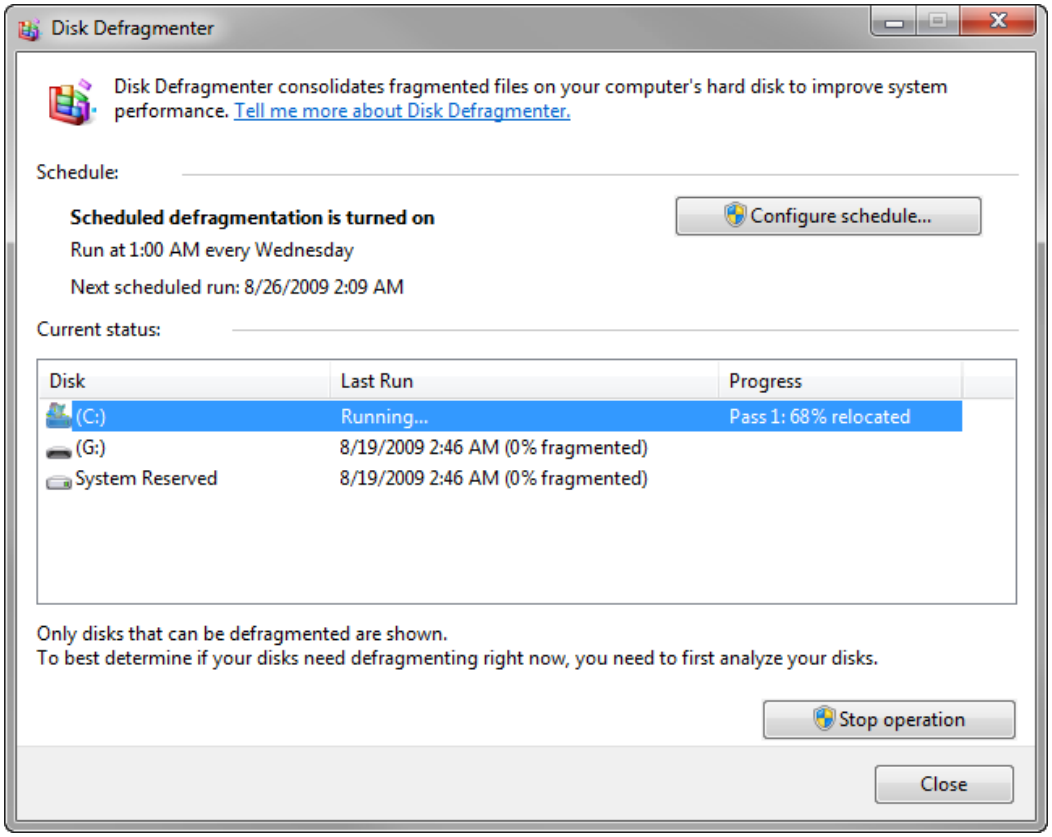

Figure 2: The UI in Windows 7's native defragmentation solution.

In this image, the user is informed that the defragmentation process is occurring, that it is running one of a series of passes, and that the process is 68% complete. Considering the performance impacts of this process that have already been discussed, you might want to provide more information to keep your users informed about the status of their defragmentation process.

The second, and more important, omission relates to the level of centralized control available to administrators. In short, Windows' native disk defragmenter has no exposure for policy-based configuration. Thus, administrators cannot create or modify an enterprise defragmentation configuration using tools such as Group Policy. Nor can administrators gain an understanding of system health across their managed computers through centralized reporting. As such, using the Windows native defragmentation solution in many ways transfers the responsibility for defragmentation away from administrators and to the user. The result is that administrators lose the ability to take action based on information gathered through any centralized information-gathering solution.

Windows Does Have This, But…

Native tools by nature enable only limited capabilities. To that end, this article series has attempted to show three things: First, that defragmentation is indeed a problem that is a naturally-occurring part of file systems operations. Second, that defragmentation is a necessary requirement of any Windows-based network. Third, the rudimentary capabilities to accomplish this process are a part of the Windows OS. However, as has been noted in this third article, they are limited in their functionality while at the same time can add a performance impact on servers and workstations.

In all of this, never forget that ultimately the sole purpose of defragmentation is to increase system performance. Save yourself the headache of freezes, crashes, and the potential for expensive purchases down the line, and consider incorporating the right kinds of defragmentation solutions into your environment.