OpenAFS for Windows Deep Dive: Reparse Points, Path Processing, and Implications for Namespace Design

Jeffrey Altman
Your File System Inc.
2014 European AFS and Kerberos Conference

First a Diversion, the Explorer Shell Caching Bug

YourFileSystem

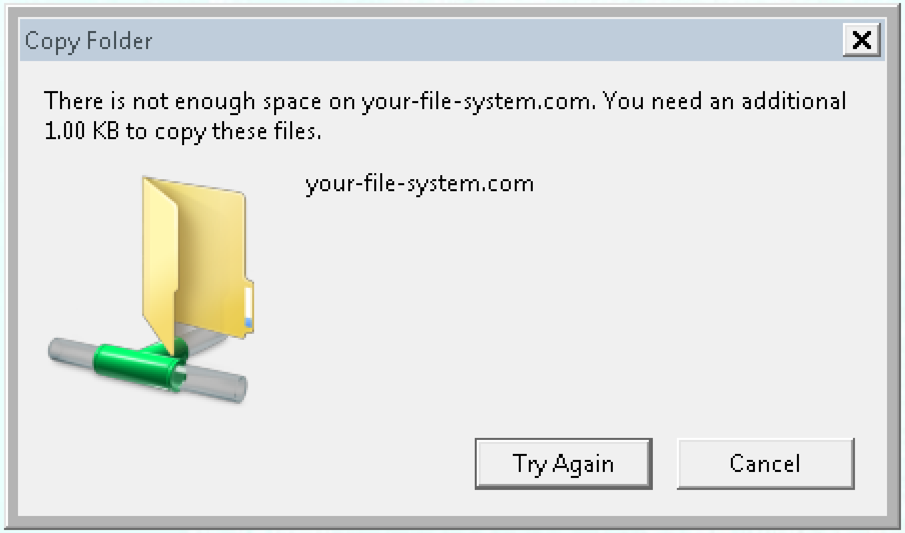
What is the Explorer Shell Caching Bug?

- The Shell caches for each directory object
  - Attributes
  - Volume Information → (AFS Volume ID = Device ID)
- The Shell caches entries for directories without attributes or volume information
- Shell believes that no Reparse Point has been crossed
- Volume Info unknown so queries info for \\afs\cell\

YourFileSystem

What are the Symptoms?

- If \\afs\cell\ refers to RO then attempt to write triggers READ ONLY VOLUME or 0 bytes free error
- If \\afs\cell refers to small RW volume, then insufficient space error is possible



Copy Folder

There is not enough space on your-file-system.com. You need an additional 1.00 KB to copy these files.

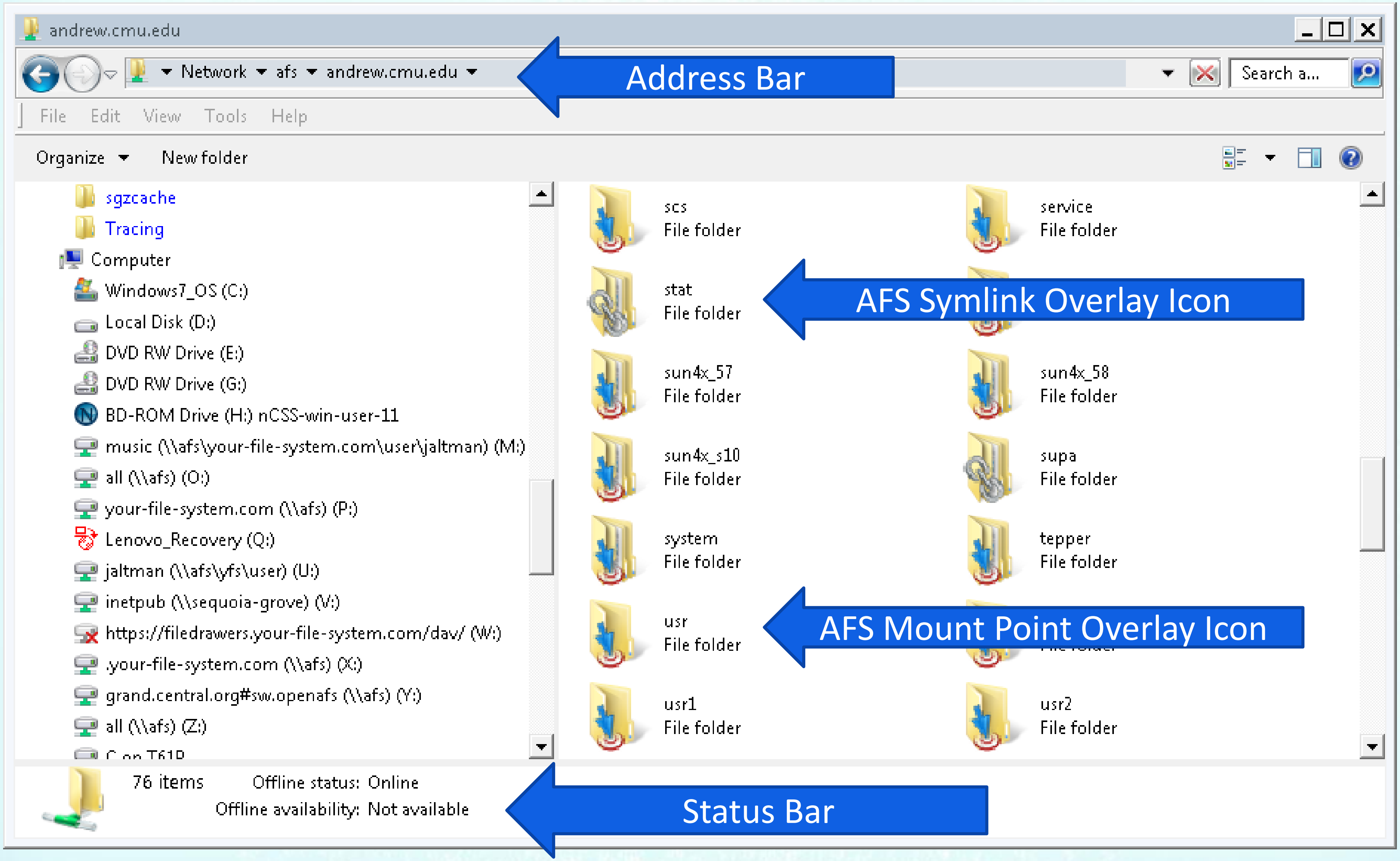
your-file-system.com

What can you do to help?

- There are open cases with Microsoft's dev team from multiple parties
- Microsoft admits this is a bug
- More reports are required to get the priority raised and a fix back-ported to at least Windows 7
- All versions of Windows are affected

YourFileSystem

The Explorer Shell



76 items Offline status: Online

Offline availability: Not available

Address Bar

AFS Symlink Overlay Icon

AFS Mount Point Overlay Icon

Status Bar

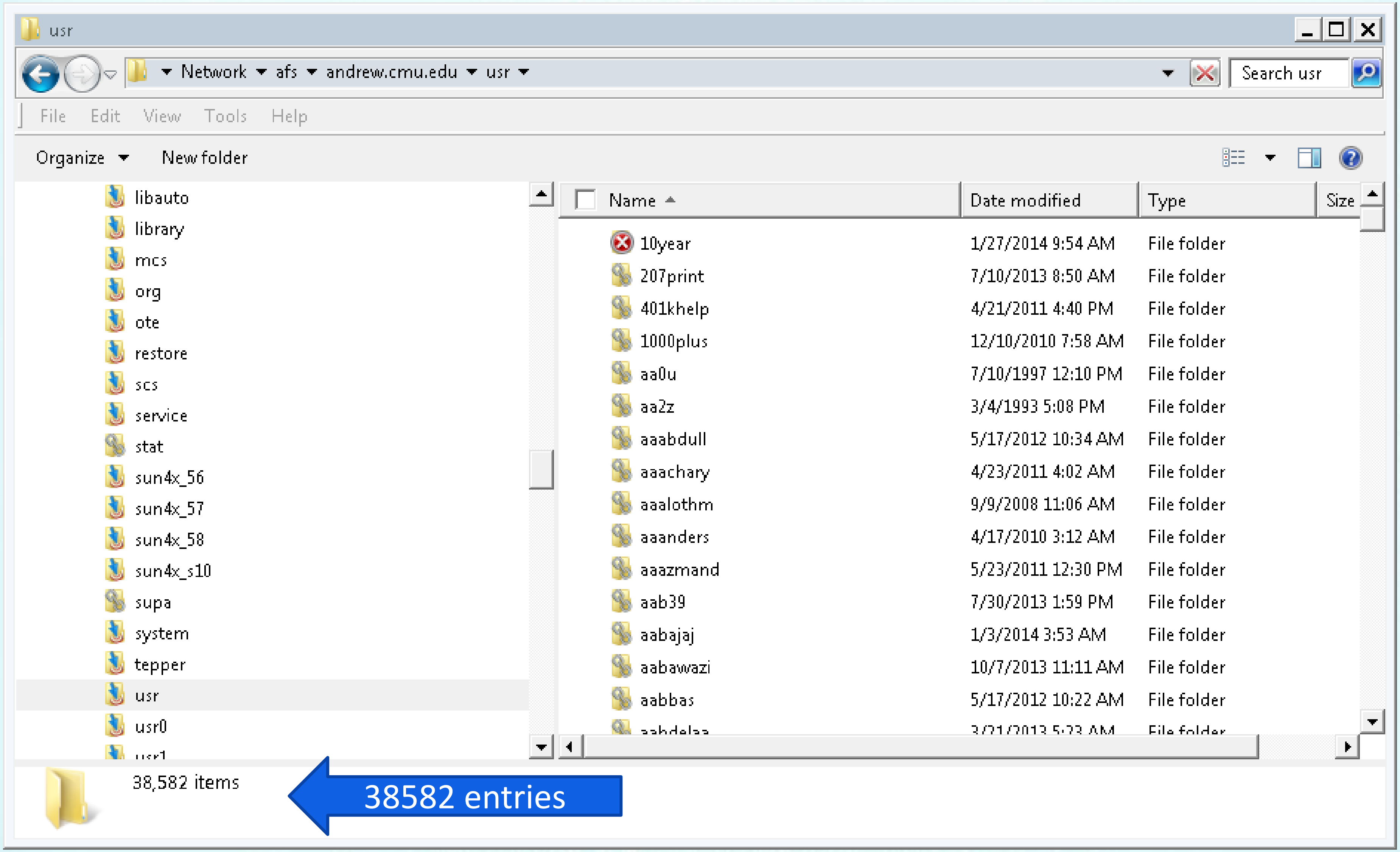
Returning to our regularly scheduled programming

YourFileSystem

Why is \\afs\andrew.cmu.edu\usr a bad idea

- The usr directory contains more than 38,000 symlink entries
  - /afs/andrew/usr/tequila -> (symlink) /afs/andrew/usr11/tequila -> (mp) #user.tequila
- Windows requires that all directory entries be presented with full status info
- All symlinks must be evaluated at least until the mount point
  - Nearly 85,000 stat objects required

/afs/andrew.cmu.edu/usr/



Name	Date modified	Type	Size
10year	1/27/2014 9:54 AM	File folder	
207print	7/10/2013 8:50 AM	File folder	
401khelp	4/21/2011 4:40 PM	File folder	
1000plus	12/10/2010 7:58 AM	File folder	
aa0u	7/10/1997 12:10 PM	File folder	
aa2z	3/4/1993 5:08 PM	File folder	
aaabdull	5/17/2012 10:34 AM	File folder	
aaachary	4/23/2011 4:02 AM	File folder	
aaalothm	9/9/2008 11:06 AM	File folder	
aaanders	4/17/2010 3:12 AM	File folder	
aaazmand	5/23/2011 12:30 PM	File folder	
aab39	7/30/2013 1:59 PM	File folder	
aabajaj	1/3/2014 3:53 AM	File folder	
aabawazi	10/7/2013 11:11 AM	File folder	
aabbas	5/17/2012 10:22 AM	File folder	
aabdelaa	3/21/2013 5:23 AM	File folder	

38,582 items

38582 entries

Status Info and Callbacks

- The default stats cache on Windows is 10,000 entries
- 85,000 entries produces large amounts of stat cache thrashing
- This is exacerbated by the AFS Redirector design that requires whole directories including status to be present in kernel

FindFirst / FindNext vs Explorer Shell

YourFileSystem

AFS Redirector vs FindFirstFile / FindNextFile / FindClose

- Current 1.7 behavior
  - Construct full directory in kernel plus status info before returning from FindFirstFile
  - Optimized to return full buffers to application
  - Results in application blocking on slow links with large directories

AFS Redirector vs FindFirstFile / FindNextFile / FindClose

- Proposed behavior
  - Construct full directory listing in service in FindFirstFile
  - Request entries plus status in FindNextFile
  - Return as many entries as possible within 200ms
  - Blocking
    - Waits for the directory enumeration in FindFirstFile
    - Waits for status info on first FindNextFile entry

Benefits of Directory Enum changes

- Explorer Shell remains responsive
- File Count increases as entries are added
- Shell Extensions are more likely to access objects while their status is in the AFSCache

AFS Symlinks vs Microsoft Symlinks

- AFS Symlinks are POSIX
  - Target type is not encoded in the target path
  - Relative or absolute paths
  - Forward slash separators
- MSFT Symlinks are not POSIX
  - Target type is encoded in the symlink object
    - Either a directory or a file with RP Data attached
  - Relative or absolute paths
  - Backslash separators

Callback Processing Changes

BEFORE

- Callback Expiration processed in the service for the afs redirector

AFTER

- AFS redirector processes its own callback expiration

Benefits:

1. File Status can be recycled in AFSCache without invalidating kernel data.
2. Fewer userland -> kernel IOCTL calls reduces CPU utilization.

Directory Enumeration Changes

BEFORE

- All directories fully populated in kernel with complete status information

AFTER

- Sparse directory enumeration in kernel
- Entries cached as needed

Benefits:

1. Fewer directory entries allocated
   1. Smaller kernel memory footprint
   2. Less CPU spent on garbage collection
   3. Fewer MPs and Symlinks evaluated
   4. Fewer RPCs issued

YourFileSystem

The Results

- 10% to 15% reduction in wall clock time when building OpenAFS Windows in /afs over WAN.
- 30% reduction in AFS Service / kernel CPU time.

When Can You Have It?

- 1.7.3001 (unofficial) build on lanyard USB disks includes these changes.   Permits comparison to 1.7.30.
- Approximately 80 patches in Gerrit
- This is a major architecture design change
  - 1.7.30 is very stable -> 1.8.0?
  - Merge these changes into new 1.9 series?
    - With or without UNIX?

More Windows Updates

- Windows 8.1 Update 1 / Server 2012 R2 Update 1
  - No changes directly affecting file systems
- OpenAFS
  - No “Modern” file browser
  - No Windows Phone client
  - No Surface RT / Surface 2 support
    - No 3rd party ARM drivers

Microsoft OneDrive in Win 8.1

- Offline file support
- Thumbnails for unavailable files
- LiveID Authentication
- Integration at Explorer Shell and “Modern” UI layers
  - Not at the file system layer
  - Third party lockout
- Must write our own File Browser for “Modern”
  - Volunteers?

YourFileSystem

A Final Word on Windows XP

- 1.7.30 is the last build that I will test on XP
- Strong desire to stop supporting Server 2003
- Vista / Server 2008 as minimum permits use of additional functionality

YourFileSystem

Questions! Answers?

Your File System

255 W 94TH ST
New York NY 10025 USA
+1 212 769-9018
sales@your-file-system.com
http://www.your-file-system.com