SAMBA in the Age of Microsoft Glasnost

ubiqx
Consulting, Inc.

Christopher R. Hertel
Storage Architect, CIFS Geek
Founder and CTO

TCCC 2011

Octoberly, 2011

Introductions

Introducing: Me

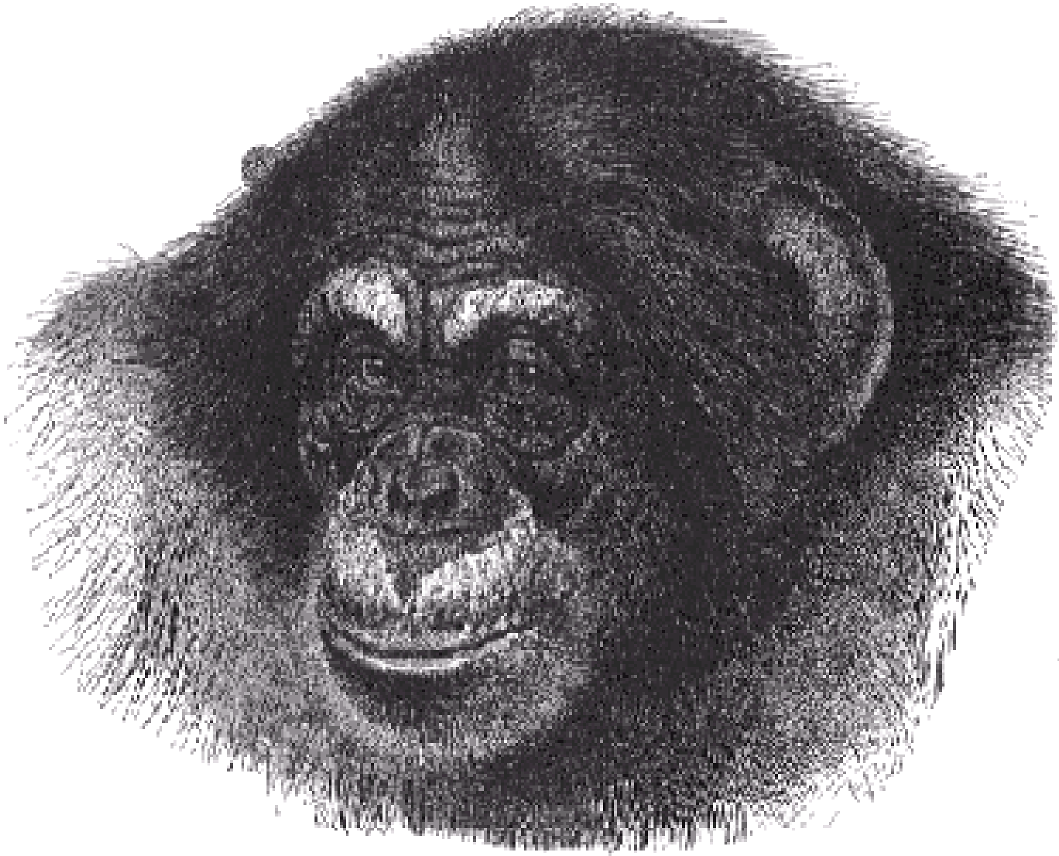
Your Friendly Neighborhood CIFS Geek

- Samba Team member (since 1998-ish)
- jCIFS Project co-founder
- CIFS Author (shameless plug)
- Network Storage Geek
- Incurable Idealist
- Etc., etc., ad nauseum

A ruminant mammal (Geekus geekus) with long legs, humped shoulders, and broadly palmated antlers.

Introducing: You

Code Monkeys!



This is Code Camp, after all.

Introducing: The Samba Team

Members of the Samba Team gather at the 10th annual Samba eXPerience conference in Göttingen, Germany.

Introducing: SMB/CIFS/SMB2

SMB = Server Message Block protocol

A stateful network file system protocol originally created by IBM in the early 1980s for use with the PC-DOS operating system.

CIFS = Common Internet File System

A “marketing upgrade” to SMB. This new name for SMB was coined in the mid 1990's. The term “CIFS” is now often used as a name for the complete suite of protocols that include and provide support for SMB. Often written “SMB/CIFS”.

SMB2 = Server Message Block protocol version 2

A complete rewrite of the SMB protocol, introduced with Windows Vista. SMB2 reduces the top-level command set from 75 commands to 19.

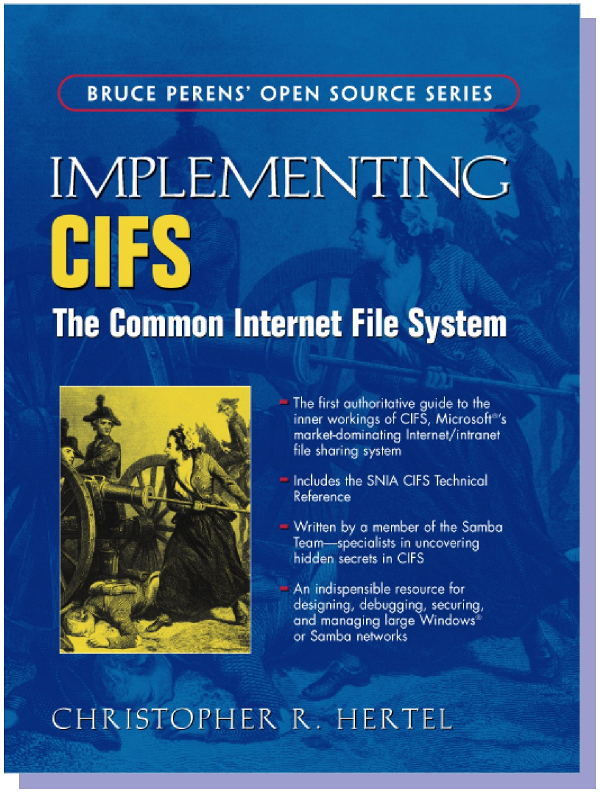
Introducing: More About Me

jCIFS

Java SMB/CIFS
Client Toolkit

Some Trouble I've Caused...

- The very best developer's guide to SMB/CIFS.
- The very worst developer's guide to SMB/CIFS.
- The very *only* developer's guide to SMB/CIFS.



BRUCE PERENS' OPEN SOURCE SERIES

IMPLEMENTING
CIFS
The Common Internet File System

- The first authoritative guide to the inner workings of CIFS, Microsoft®'s market-dominating Internet/intranet file sharing system
- Includes the SNIA CIFS Technical Reference
- Written by a member of the Samba Team—specialists in uncovering hidden secrets in CIFS
- An indispensible resource for designing, debugging, securing, and managing large Windows® or Samba networks

CHRISTOPHER R. HERTEL

Me Me Me

Open Source Credentials Notwithstanding...

Microsoft asked a member of the *Samba Team* to document SMB/CIFS!

Introducing: The Docs

Thus, SMB/CIFS is covered in two documents:

[MS-CIFS]

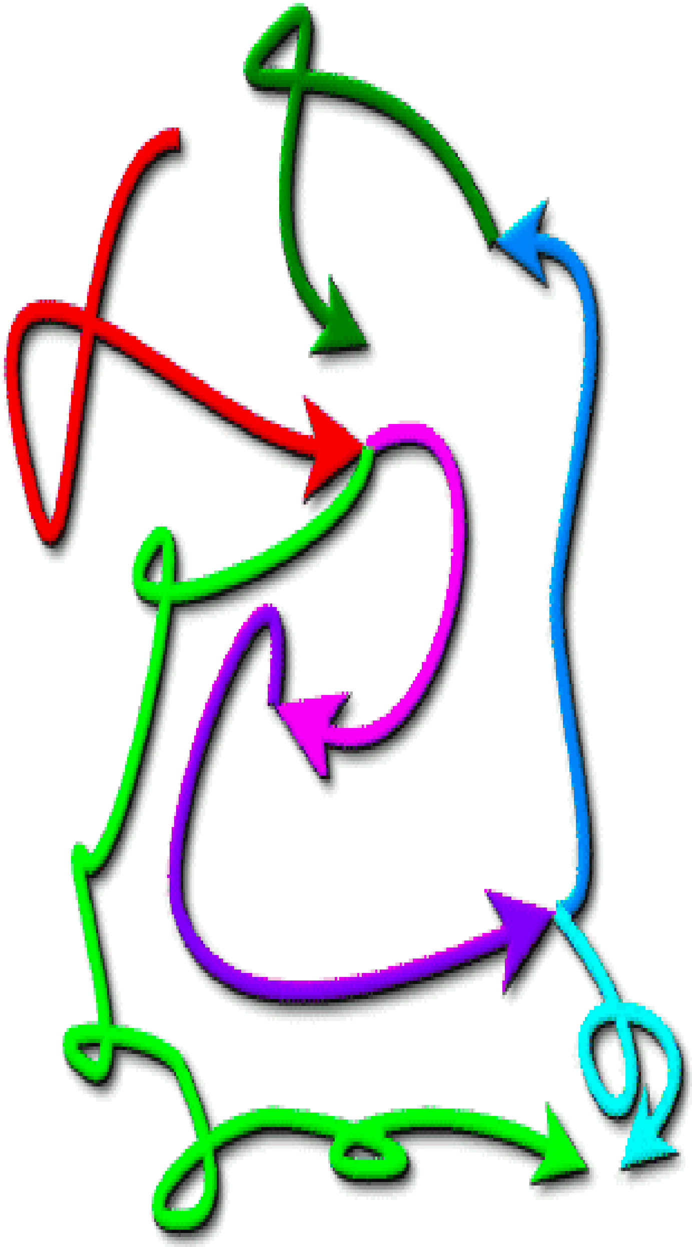
- Provides the base specification of the "NT LM 0.12" dialect.
- A "snapshot in time".
- Most of this stuff is still there in current Windows versions. Really.

[MS-SMB]

- "Extends" [MS-CIFS].
- Documents changes made to SMB starting in W2K.
- Still the same "NT LM 0.12" dialect.

Note: The naming is backwards!

Where are we going?

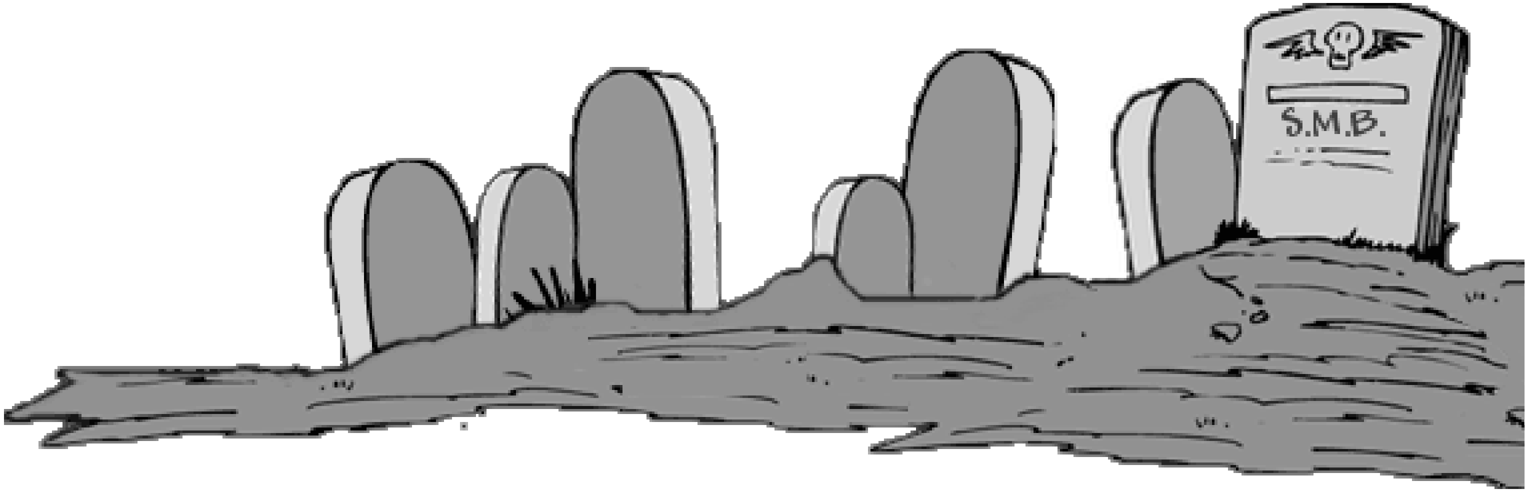


- SMB/CIFS is Dead
  - Long live SMB2?
- The Entourage
  - 400 documents, give or take
- BITS Upload Protocol
  - An easy start
- PeerDist Protocol
  - Three-part harmony
- SMB2.2
  - The future's so bright, I gotta wear shades

CIFS is Dead

CIFS is Dead

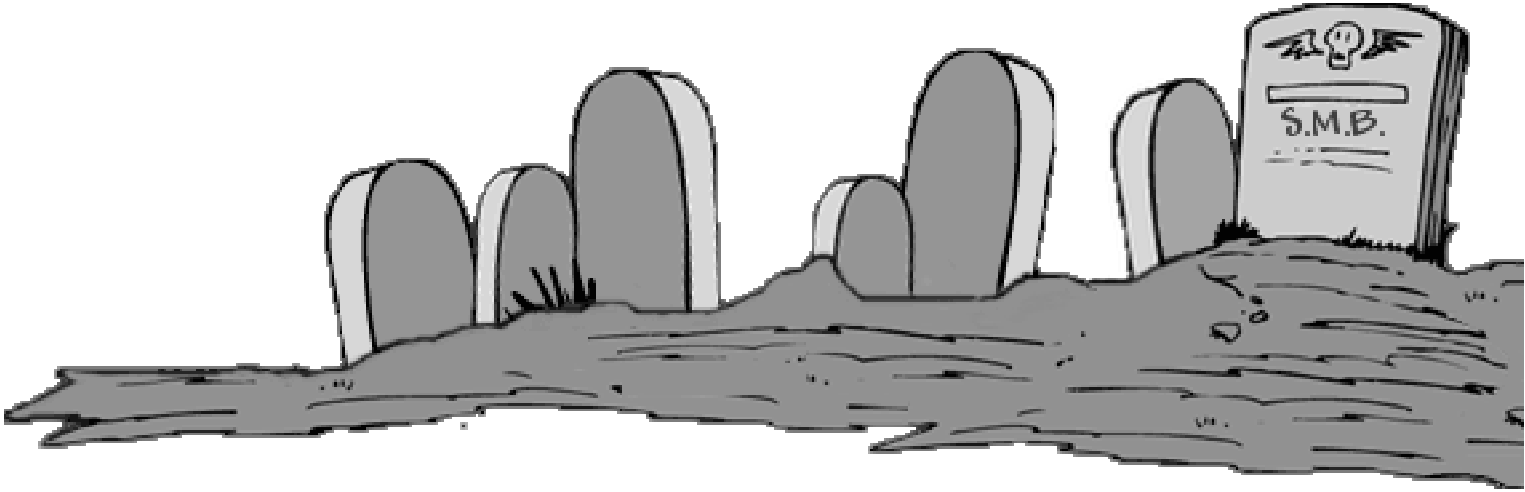
- The current dialect is “NT LM 0.12”.
  It was introduced with Windows NT.



CIFS is Dead

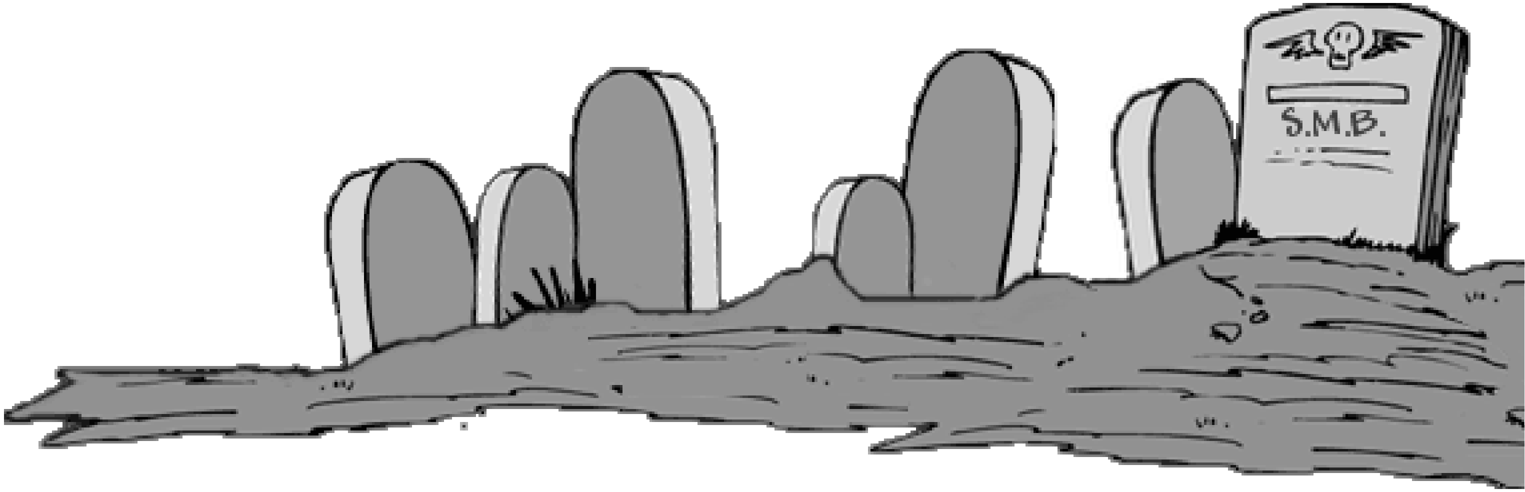
- The current dialect is “NT LM 0.12”.
  It was introduced with Windows NT.

- New dialects of SMB/CIFS are unlikely.



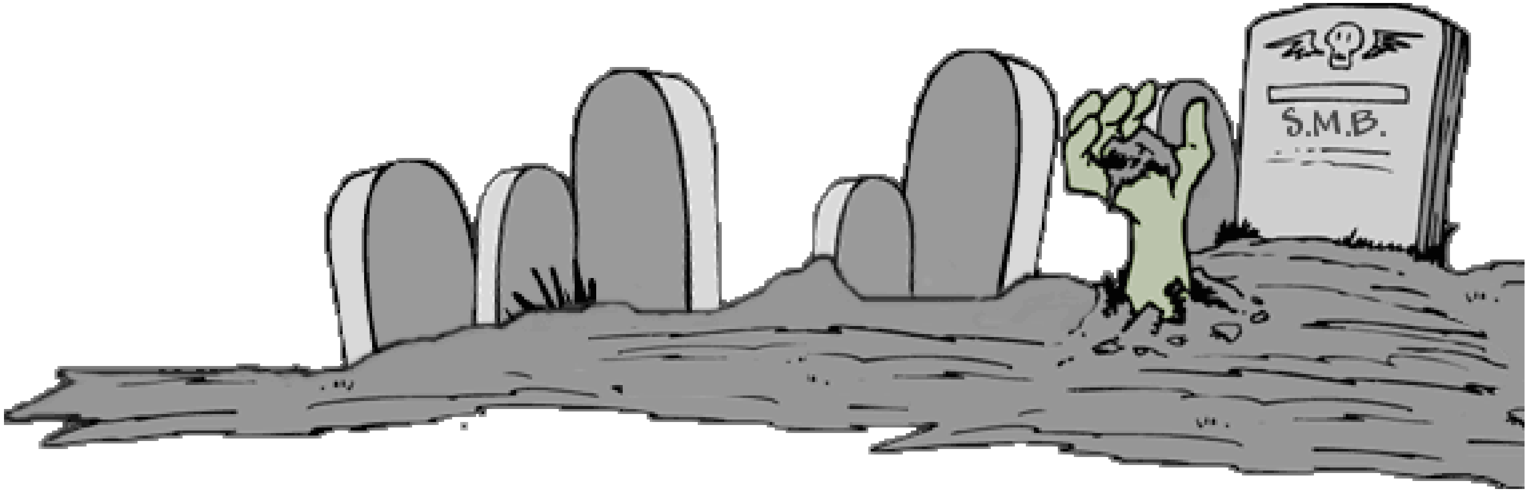
CIFS is Dead

- The current dialect is “NT LM 0.12”.
  It was introduced with Windows NT.
- New dialects of SMB/CIFS are unlikely.
- Microsoft is focused on SMB2.



CIFS is Undead

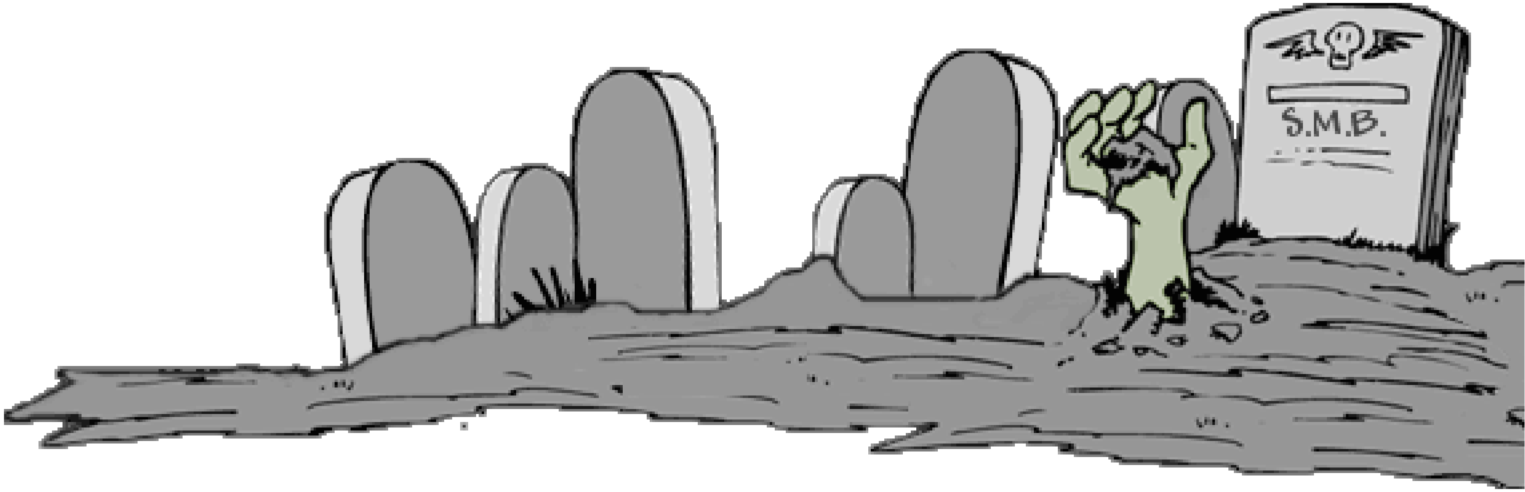
- SMB/CIFS is still the most widely used network file system on the planet, by a wide margin.
- Supported in all versions of MS-Windows.
- Supported by 3rd-party NAS Vendors.



CIFS is Undead

SMB/CIFS code examples:

- Samba, of course
- jCIFS – Smaller, simpler, client-only
- Sample code in Implementing CIFS
- See also: http://www.ubiqx.org/libcifs/



CIFS is *Undead* and Living at:

CIFS.ORG

The Docs

The Docs

Go here:
http://www.microsoft.com/openspecifications/

Over 400 documents have been published, covering:

- Authentication
- Windows Internals
- File Formats
- Client-Server Protocols
- Server-Server Protocols

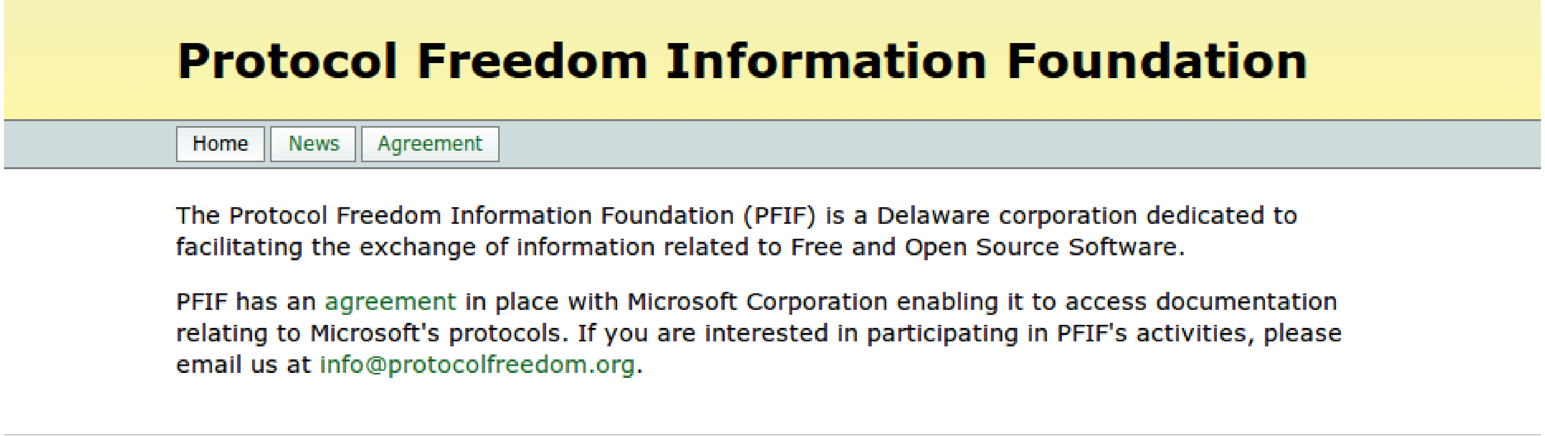
Overview docs provide starting points for understanding groups of docs.

The Docs

"We should implement them all."

- There is an opportunity here to leverage both the technology and the installed base.
- The documentation includes previews of SMB2.2, and other features of Windows 8.
- This will feed the software engineering ecosystem for years.

The Docs



Protocol Freedom Information Foundation

The Protocol Freedom Information Foundation (PFIF) is a Delaware corporation dedicated to facilitating the exchange of information related to Free and Open Source Software.

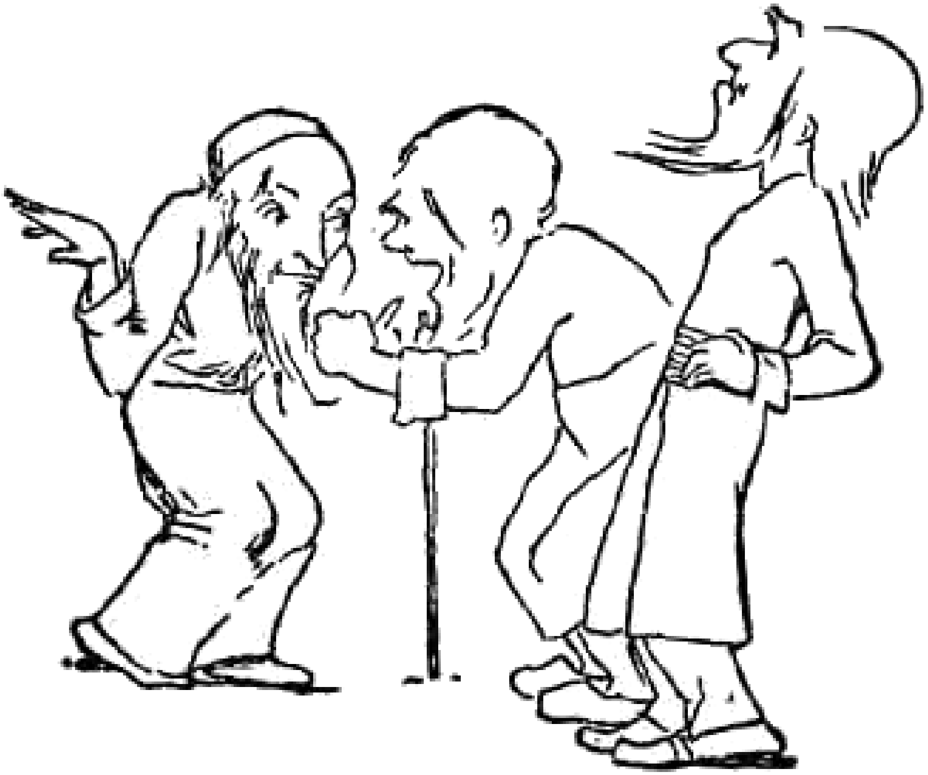
PFIF has an agreement in place with Microsoft Corporation enabling it to access documentation relating to Microsoft's protocols. If you are interested in participating in PFIF's activities, please email us at info@protocolfreedom.org.

- The PFIF provides subcontractors with access to Microsoft documentation under the terms of the Work Group Server Protocol Program (WSPP).
- See the PFIF website for more information.
- See also the “Open Specification Promise”.

The Docs

Okay...

...so what can you *do* with all of this?



BITS

01000010 01100001
01100011 01101011
01100111 01110010
01101111 01110101
01101110 01100100
00100000 01001001
01101110 01110100
01100101 01101100
01101100 01101001
01100111 01100101
01101110 01110100

BITS: Background “Intelligent” Transfer Service

“BITS is Earth’s most widely used file transfer service, with more than 600 million unique users across the planet.”

– Vipul Bansal, Microsoft WMI Blog, Jan 2009.

BITS: Background “Intelligent” Transfer Service

“BITS is Earth’s most widely used file transfer service, with more than 600 million unique users across the planet.”

– Vipul Bansal, Microsoft WMI Blog, Jan 2009.

Note Well:

BITS: Background "Intelligent" Transfer Service

"BITS is Earth's most widely used file transfer service, with more than 600 million unique users across the planet."

– Vipul Bansal, Microsoft WMI Blog, Jan 2009.

Note Well: *nobody cares.*

BITS: Background “Intelligent” Transfer Service

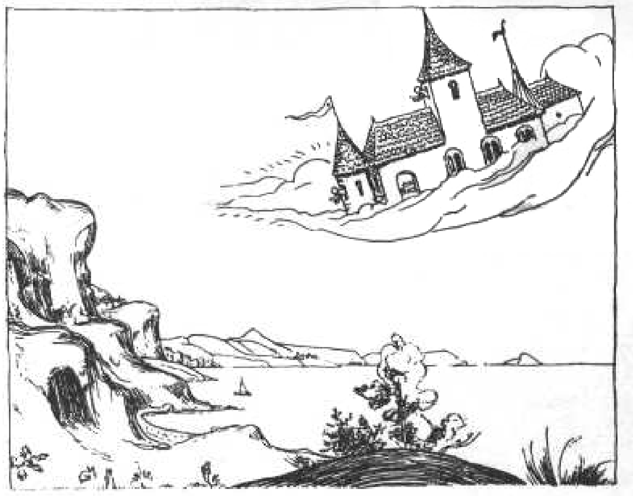
“BITS is Earth’s most widely used file transfer service, with more than 600 million unique users across the planet.”
– Vipul Bansal, Microsoft WMI Blog, Jan 2009.

What does that mean anyway?

- It does not say “protocol”, it says “file transfer service”.
- BITS is the Windows *system service* used by Windows Update to download patches.
- Most users don't even know it's there.

BITS: Background "Intelligent" Transfer Service

BITS Features



- Built into Windows
- Restartable Transfers
  - ...but only linearly; does not "patch".
- Both Download and Upload
  - ...and "Upload Reply".
- Job priority levels
- Senses network traffic to manage impact

BITS: Background “Intelligent” Transfer Service

BITS Download Jobs

- The overwhelming majority of BITS jobs are probably Windows Update downloads.
- BITS Downloads use HTTP/HTTPS.
- Sort of like `uucp`?
  `wget` + `batch` + `nice` + `diffserv`?

The “special sauce” is the use of network traffic monitoring to limit BITS data transfer rates.

BITS: Background “Intelligent” Transfer Service

BITS *Up*load Jobs

- Much less common.
- Proprietary extensions to HTTP/HTTPS.
- Only between Windows BITS clients and Windows HTTP[S] servers.

BITS: Background “Intelligent” Transfer Service

BITS *Up*load Jobs

- Much less common.
- Proprietary extensions to HTTP/HTTPS.
- Only between Windows BITS clients and Windows HTTP[S] servers – Until now!

BITS: Background "Intelligent" Transfer Service

STiB means:

- **S**low **T**ransfer **in** **B**ackground?
- **S**illy **T**echnology **is** **B**oring?
- **S**ipping **T**ea **in** **B**elgium?
- BITS spelled sdrawkcab with a small '*i*'?

STiB: It Is what It Is.

- ...a toolkit for testing BITS Uploads.
- ...example code for others to read / use.

A CGI script could be written to
accept BITS Uploads.

BITS: Background “Intelligent” Transfer Service

BITS Upload Extensions:

- HTTP Extension Method: BITS_POST
- BITS Packet Types
  - Ping
  - Create-Session
  - Fragment
  - Cancel-Session
  - Close-Session
  - Ack

BITS Documentation:

- MSDN: BITS Upload Protocol
- WSPP: [MC-BUP]

BITS: Background “Intelligent” Transfer Service

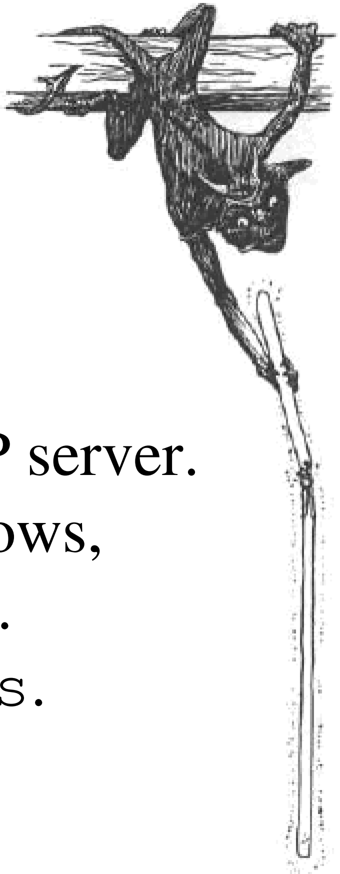
BITS Upload Extensions:

- HTTP Extension Method: BITS_POST
- BITS Packet Types
  - Ping
  - Create-Session
  - Fragment
  - Cancel-Session
  - Close-Session
  - Ack

BITS Documentation:

- MSDN: BITS Upload Protocol
- WSPP: [MC-BUP]

BITS: Background "Intelligent" Transfer Service

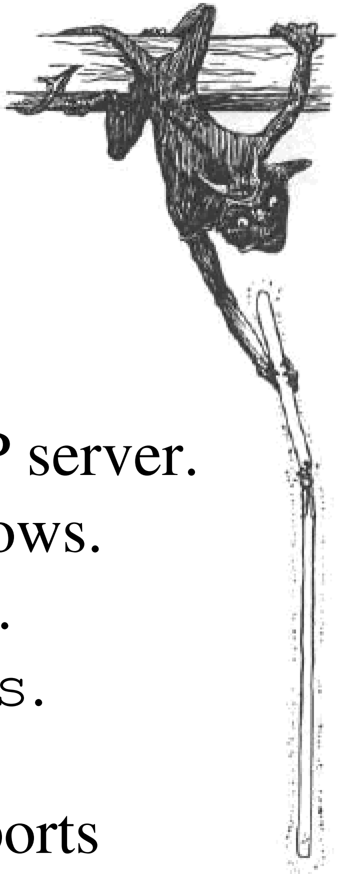


Do we care?

YAWP (Yet Another Windows Protocol)

- BITS Upload is supported in IIS,
  - and in Microsoft's "lightweight" HTTP server.
- It's convenient when working with Windows,
  - Not nearly as powerful as, eg., `rsync`.
  - Not as secure as `sftp`, `scp`, or `sshfs`.

BITS: Background “Intelligent” Transfer Service



Do we care?

YAWP (Yet Another Windows Protocol)

- BITS Upload is supported in IIS,
  - and in Microsoft's “lightweight” HTTP server.
- It's convenient when working with Windows.
  - Not nearly as powerful as, eg., `rsync`.
  - Not as secure as `sftp`, `scp`, or `sshfs`.

MS-BITS, however, also supports BranchCache™, which suggests some very useful testing scenarios.

- `GET` support added to STiB,
- PeerDist included in the header,
- It works!

BITS: Background “Intelligent” Transfer Service

http://www.ubiqx.org/proj/STiB/

STiB is at version 0.2

It contains **`stibtest`**, which can:

- Send files using MS-BUP protocol,
- Get files using HTTP1.1,
  - Get a subrange of a requested file,
- Specify “peerdist” encoding when requesting all or part of a file.

Please download and test it.
Send patches.

BITS: Background “Intelligent” Transfer Service

http://www.ubiqx.org/proj/STiB/

“Someone” should write an Apache Module that to handle BITS Upload Protocol.

Prequel

Pay Attention!

This is where it gets interesting.

Prequel

What the heck is *Prequel*?

Prequel

Prequel: A project to build an
Open Source Implementation
of Microsoft's BranchCache™.

So what the heck is BranchCache™?

Prequel

Prequel: A project to build an Open Source Implementation of Microsoft's BranchCache™.

BranchCache™ is a distributed content caching system

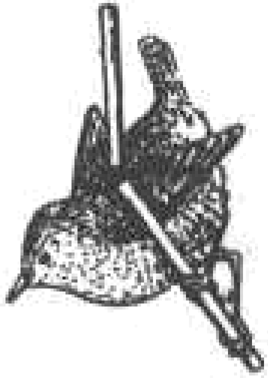
- supported in W2K8r2 servers,
- and Windows7 clients.

Cheap, effective WAN acceleration for SMB2, HTTP, and BITS.

Prequel

BranchCache Architecture

A quick overview



Content Servers

- Have content to share with multiple clients.

Clients (peers)

- Request & receive content from content servers.

The Cache

- A copy of the original content, divided into segments and blocks, accessed via hash tags.

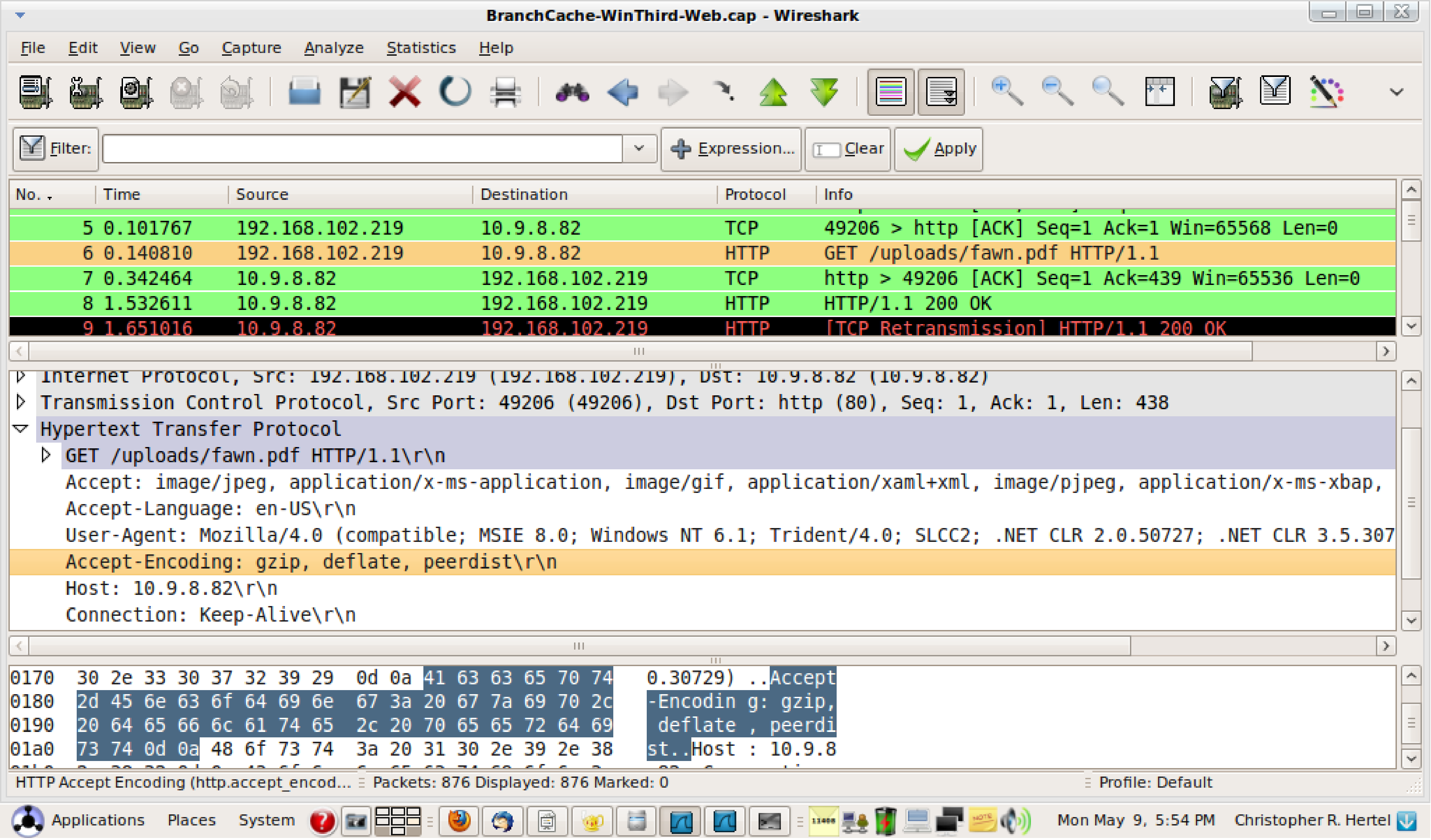
Prequel

Content Servers:

- Web Servers (HTTP, BITS)
- File Servers (SMB2.1)

The client must know to ask for *content tags* instead of actual content.

- If the tags are already calculated, they are returned by the BranchCache™-enabled server.
- Otherwise the content is returned, and the server (W2K8r2) calculates the tags for next time.



```
No. .   Time       Source            Destination       Protocol  Info
  5 0.101767      192.168.102.219   10.9.8.82         TCP       49206 > http [ACK] Seq=1 Ack=1 Win=65568 Len=0
  6 0.140810      192.168.102.219   10.9.8.82         HTTP      GET /uploads/fawn.pdf HTTP/1.1
  7 0.342464      10.9.8.82         192.168.102.219   TCP       http > 49206 [ACK] Seq=1 Ack=439 Win=65536 Len=0
  8 1.532611      10.9.8.82         192.168.102.219   HTTP      HTTP/1.1 200 OK
  9 1.651016      10.9.8.82         192.168.102.219   HTTP      [TCP Retransmission] HTTP/1.1 200 OK
```

```
▷ Internet Protocol, Src: 192.168.102.219 (192.168.102.219), Dst: 10.9.8.82 (10.9.8.82)
▷ Transmission Control Protocol, Src Port: 49206 (49206), Dst Port: http (80), Seq: 1, Ack: 1, Len: 438
▽ Hypertext Transfer Protocol
  ▷ GET /uploads/fawn.pdf HTTP/1.1\r\n
    Accept: image/jpeg, application/x-ms-application, image/gif, application/xaml+xml, image/pjpeg, application/x-ms-xbap,
    Accept-Language: en-US\r\n
    User-Agent: Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.1; Trident/4.0; SLCC2; .NET CLR 2.0.50727; .NET CLR 3.5.307
    Accept-Encoding: gzip, deflate, peerdist\r\n
    Host: 10.9.8.82\r\n
    Connection: Keep-Alive\r\n
```

```
0170  30 2e 33 30 37 32 39 29  0d 0a 41 63 63 65 70 74   0.30729) ..Accept
0180  2d 45 6e 63 6f 64 69 6e  67 3a 20 67 7a 69 70 2c   -Encodin g: gzip,
0190  20 64 65 66 6c 61 74 65  2c 20 70 65 65 72 64 69    deflate , peerdi
01a0  73 74 0d 0a 48 6f 73 74  3a 20 31 30 2e 39 2e 38   st..Host : 10.9.8
```

```
HTTP Accept Encoding (http.accept_encod...   Packets: 876 Displayed: 876 Marked: 0   Profile: Default
```

This is IE 8 indicating support for BranchCache™
by listing "peerdist" as an acceptable encoding.

`Accept-Encoding: gzip, deflate, `**`peerdist`**`\r\n`

Prequel

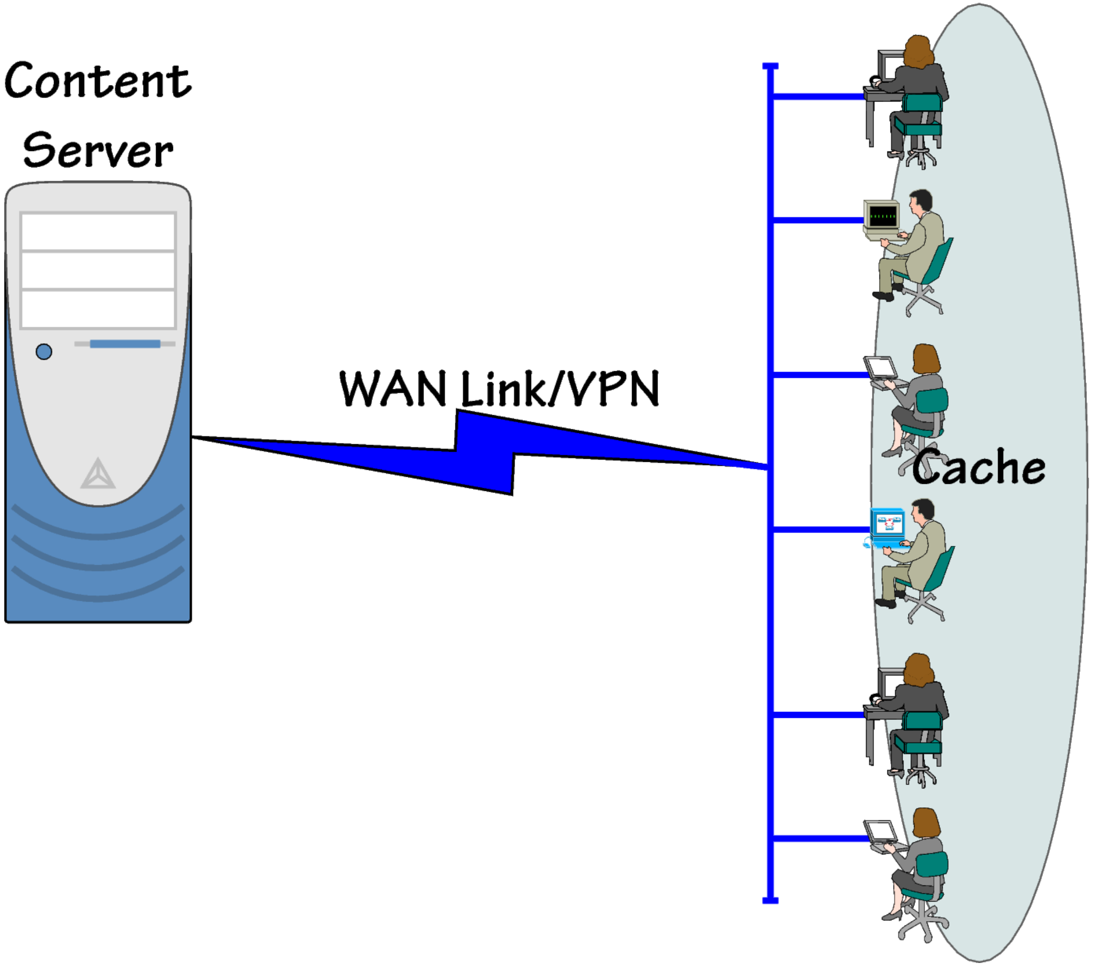
Client-side PeerDist Caching

There are two modes of operation:

- Distributed Mode
- Hosted Mode

Prequel

Distributed Mode



Prequel

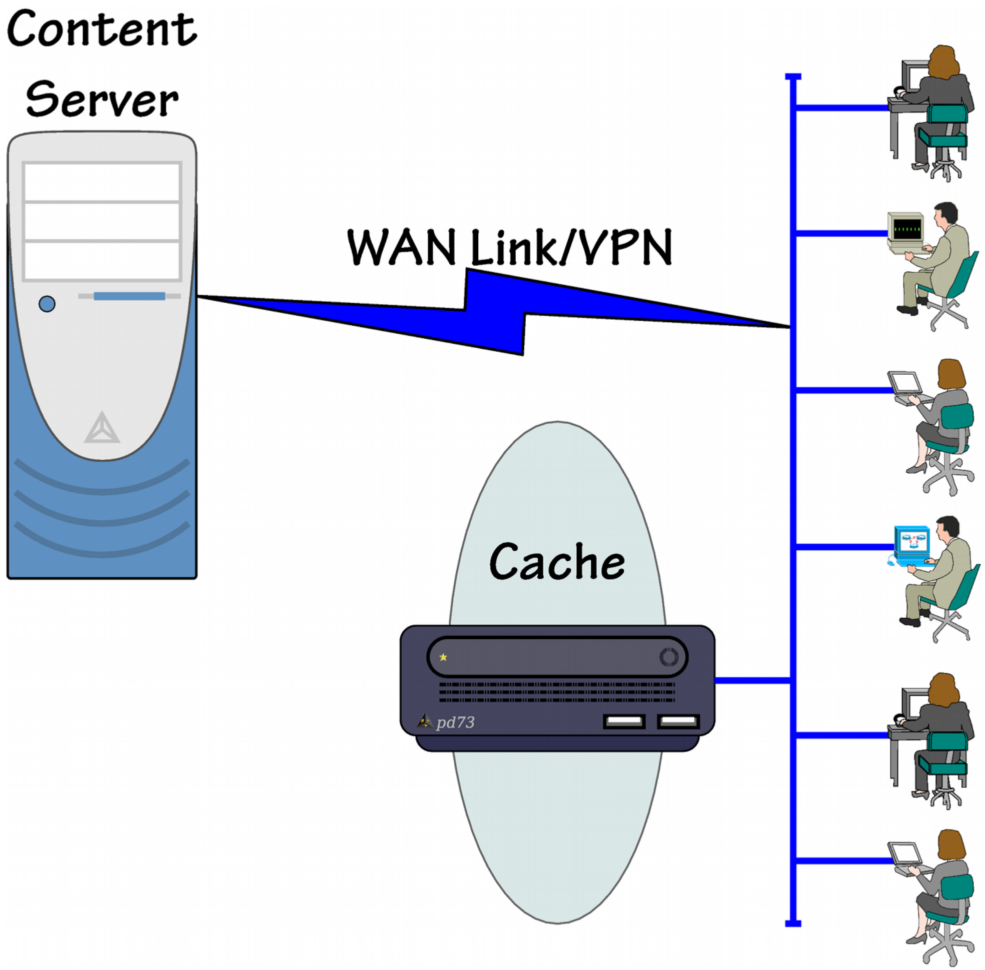
Distributed Mode

- Each client keeps a local cache.
- A client requests PeerDist tags from the server, then broadcasts to find the cached content.
- If the content is not cached,
  - The client requests the content from the content server,
  - The client stores both content and tags in its own cache.

Reminiscent of the CIFS Browse Service.

Prequel

Hosted Mode



Prequel

Hosted Mode

- A client request tags from the content server
- The client then asks the local cache server for the content
- If the content is not cached, the client requests content from the content server
- The client sends both content and tags to the cache server
- Content can now be retrieved from the cache server using only tags

Prequel

Content Tags

Blocks

- Are a unit of download
  (from either the content server or cache server)
- Are 64K
  (or less, for the last block in a file only)

The block tag is an SHA$\{\substack{256\\384\\512}\}$ hash of the block.

Segments

- Are a unit of discovery
- One segment is 32M == 512 blocks
  (or less, if the last block is short)

Segments are identified by a hash of the block hashes.

Prequel

http://www.ubiqx.org/proj/Prequel/

Prequel does not have a release number yet.

`pq_cgi` – CGI program to generate PeerDist Content Information.
- Tested with Apache.

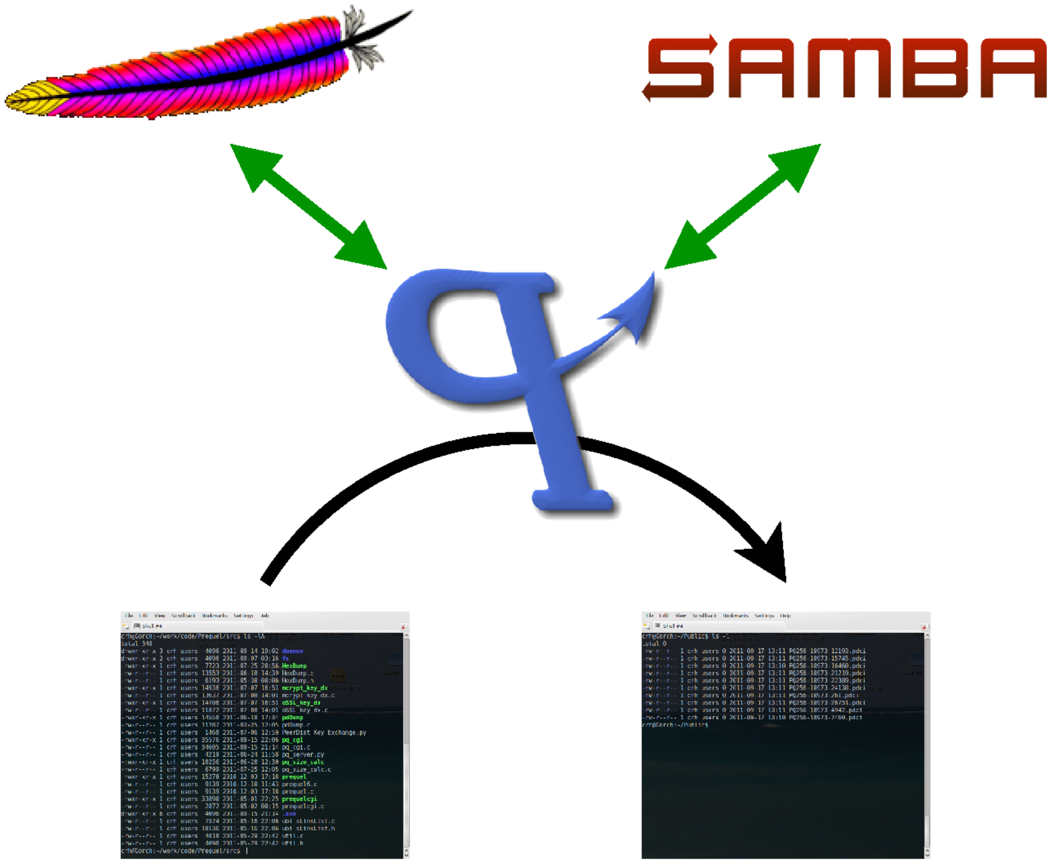
`pdDump` – Pretty-print Content Information.

`*_key_dx` – Extract W2K8r2 Server Passphrase and Server Secret

Prequel

Prequel Dæmon

Conceptual Overview


SAMBA
File Edit View Scrollback Bookmarks Settings Help
Shell #4
Source Files
File Edit View Scrollback Bookmarks Settings Help
Shell #4
Cache Files

```
crh@Gorch:~/work/code/Prequel/src$ ls -lA
total 348
drwxr-xr-x 3 crh users  4096 2011-09-14 19:02 daemon
drwxr-xr-x 2 crh users  4096 2011-09-07 03:16 fs
-rwxr-xr-x 1 crh users  7723 2011-07-25 20:56 HexDump
-rw-r--r-- 1 crh users 13553 2011-06-10 14:39 HexDump.c
-rw-r--r-- 1 crh users  6193 2011-05-30 08:06 HexDump.h
-rwxr-xr-x 1 crh users 14036 2011-07-07 16:51 mcrypt_key_dx
-rw-r--r-- 1 crh users 13637 2011-07-08 14:01 mcrypt_key_dx.c
-rwxr-xr-x 1 crh users 14708 2011-07-07 18:51 oSSL_key_dx
-rw-r--r-- 1 crh users 11872 2011-07-08 14:01 oSSL_key_dx.c
-rwxr-xr-x 1 crh users 14558 2011-06-18 17:34 pdDump
-rw-r--r-- 1 crh users 11287 2011-07-25 12:05 pdDump.c
-rw-r--r-- 1 crh users  1468 2011-07-06 12:59 PeerDist Key Exchange.py
-rwxr-xr-x 1 crh users 35576 2011-09-15 22:06 pq_cgi
-rw-r--r-- 1 crh users 34605 2011-09-15 21:14 pq_cgi.c
-rw-r--r-- 1 crh users  4219 2011-06-24 11:58 pq_server.py
-rwxr-xr-x 1 crh users 10256 2011-06-20 12:30 pq_size_calc
-rw-r--r-- 1 crh users  6799 2011-07-25 12:05 pq_size_calc.c
-rwxr-xr-x 1 crh users 15270 2010-12-03 17:10 prequel
-rw-r--r-- 1 crh users  9139 2010-12-10 11:43 prequel6.c
-rw-r--r-- 1 crh users  9139 2010-12-03 17:10 prequel.c
-rwxr-xr-x 1 crh users 33898 2011-05-01 22:25 prequelcgi
-rw-r--r-- 1 crh users  2872 2011-05-02 08:15 prequelcgi.c
drwxr-xr-x 6 crh users  4096 2011-09-15 21:14 .svn
-rw-r--r-- 1 crh users  7574 2011-05-16 22:08 ubi_sLinkList.c
-rw-r--r-- 1 crh users 10136 2011-05-16 22:06 ubi_sLinkList.h
-rw-r--r-- 1 crh users  4818 2011-05-29 22:42 util.c
-rw-r--r-- 1 crh users  4098 2011-05-29 22:42 util.h
crh@Gorch:~/work/code/Prequel/src$ |
```

```
crh@Gorch:~/Public$ ls -l
total 0
-rw-r--r-- 1 crh users 0 2011-09-17 13:11 PQ256-18973-12193.pdci
-rw-r--r-- 1 crh users 0 2011-09-17 13:11 PQ256-18973-15745.pdci
-rw-r--r-- 1 crh users 0 2011-09-17 13:10 PQ256-18973-16460.pdci
-rw-r--r-- 1 crh users 0 2011-09-17 13:11 PQ256-18973-21219.pdci
-rw-r--r-- 1 crh users 0 2011-09-17 13:11 PQ256-18973-22389.pdci
-rw-r--r-- 1 crh users 0 2011-09-17 13:11 PQ256-18973-24138.pdci
-rw-r--r-- 1 crh users 0 2011-09-17 13:11 PQ256-18973-261.pdci
-rw-r--r-- 1 crh users 0 2011-09-17 13:11 PQ256-18973-28751.pdci
-rw-r--r-- 1 crh users 0 2011-09-17 13:11 PQ256-18973-4942.pdci
-rw-r--r-- 1 crh users 0 2011-09-17 13:10 PQ256-18973-7460.pdci
crh@Gorch:~/Public$
```

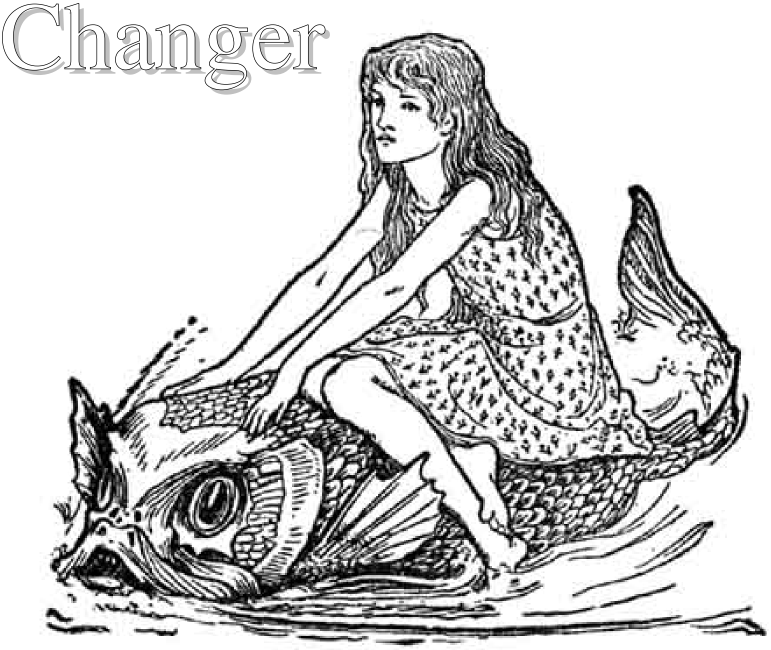
SMB2.2

SMB2.2

The Future's So Bright, I Gotta Wear Shades

SMB2.2

SMB2.2 is a Game-Changer



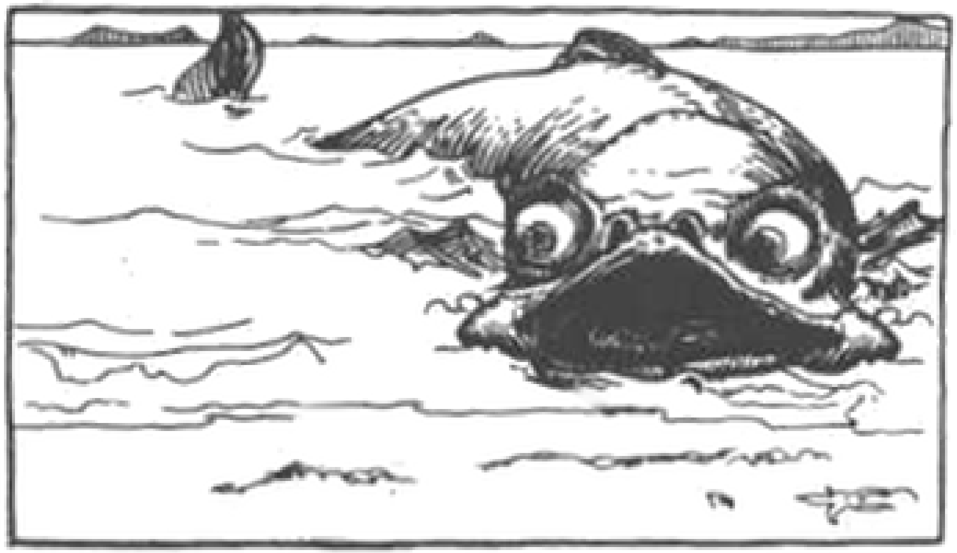
- Multipath
- Cluster capabilities
- Distributed caching
- SMB over RDMA

SMB2.2 is aimed at supporting Application Servers in Enterprise Datacenters.

It presents a strong challenge to NFSv4.x.

SMB2.2

Stocking the SMB/CIFS Pond



- The SMB/CIFS/SMB2 Talent Pool is very shallow.
- The SMB/CIFS/SMB2-talented are big fish.
- The number of organizations in the SNIA CIFS Plugfest has grown from 14 to 27 in 4 years.
- SMB2.2 is likely to generate a lot of work.

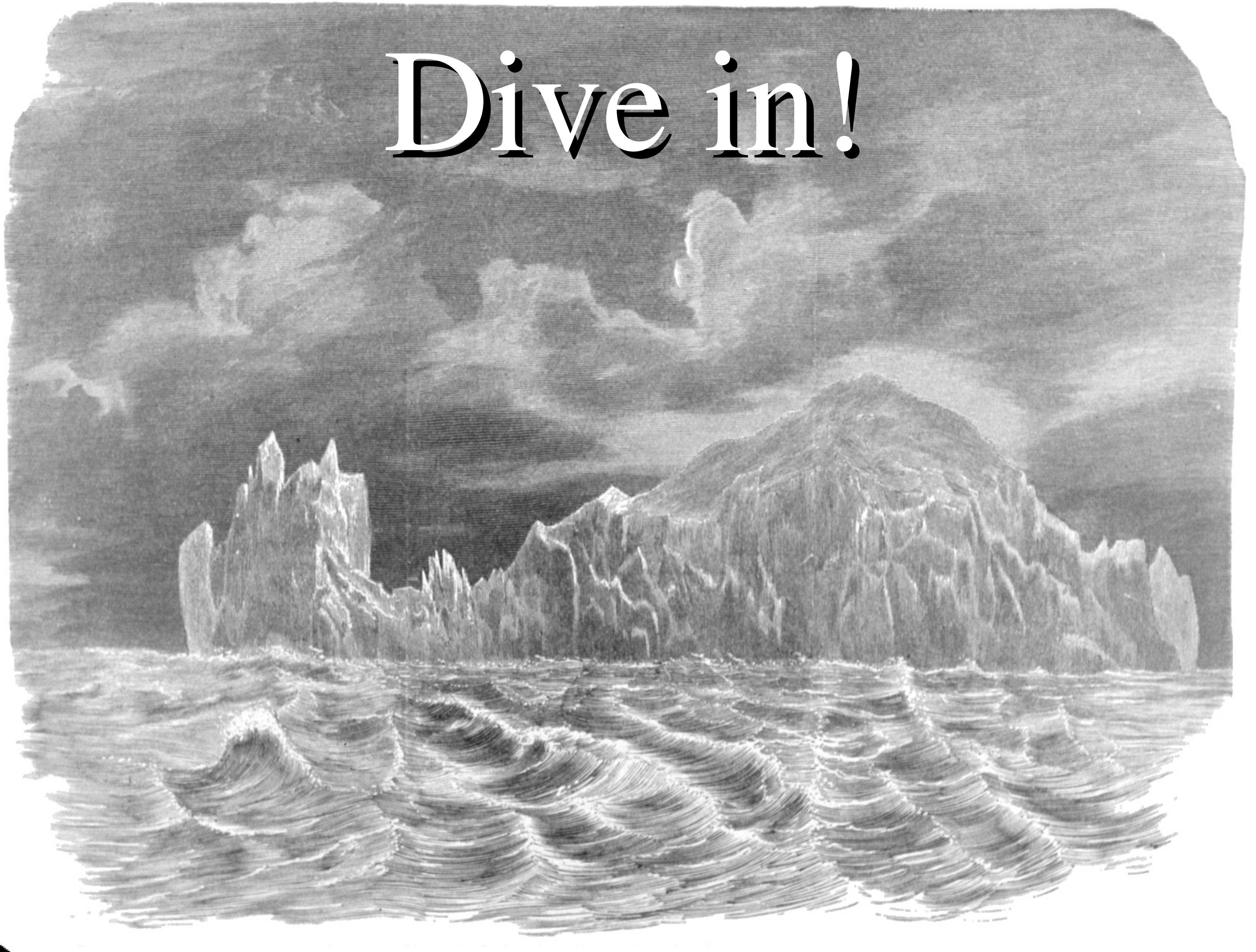
SMB2.2

Whither Samba?

- Now that Microsoft is sharing, do we need the Rebel Alliance?
- When will we see Samba 4?
- What have the Samba Team been doing for the past few years?

- Directory Services
- CTDB Clustering
- Python Bindings
- Restructure the Code

SMB2.2



The End