



GNU Wget2



About GNU Wget

- Command Line Program, first released in 1996. Developed by Hrvoje Nikšić
- Fetches files off the internet over HTTP(S) and FTP(S)
- Entirely non-interactive, ideal for shell scripts
- *Recursive Downloading*
- Continue broken downloads (HTTP/1.1)
- So stable, most think it has no developers!



The Need For Wget2

- We want to support newer web technologies.
- Lack of a good regression test suite (Unit + Functional Tests)
- New features almost impossible to implement
- Single Threaded Design
 - Blocking sockets
 - Static / Global Variables
 - Non-reentrant code
- Confusing switches



A Brief History of Wget2

- **2013:** libmget is created as a reusable library. `mget` is a command line application and example implementation of libmget API
- **2014:** Split LibPSL into own library
- **2015:** Support for HTTP/2 added via `nghttp` library
- **2015:** Mget is renamed to Wget2 and brought into GNU
- **2016:** Large portability push via gnuilib. Now supports Linux, Solaris, OSX, BSD and MingW64
- **2017:** Full CI and Fuzz Testing integration
- **2017:** Accepted three students via GSoC for new features



New and Interesting Features

- Built around an LGPL library, libwget
- (Almost) fully backwards compatible with Wget switches
- Continuous Integration and testing via GitLab CI
- Regular Static Analysis via Coverity
- Testing using Address and Undefined Behaviour Sanitizers
- Fuzz Testing via OSS-Fuzz Project
- Test Suite written entirely in C with no scripting required



New and Interesting Features

- Non-Blocking Sockets
- HTTP/2 Support
- Multi-threaded Downloading
- HTTP Compression (gzip, bz2, xz, lzma, brotli)
- TCP Fast Open (-1 RTT)
- TLS Session Resumption | TLS False Start (-1 RTT)
- XDG Base Directory Compliant
- Metalink Support

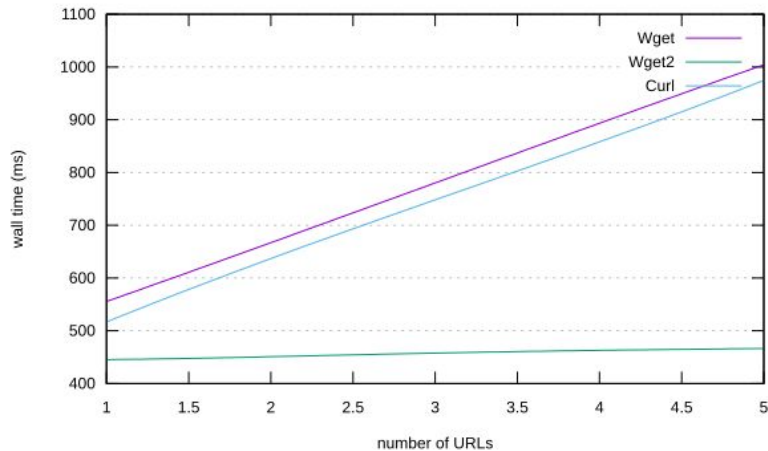


New and Interesting Features

- Uses LibPSL to test for Public Suffixes before accepting cookies
- HTTP Strict Transport Security - HSTS (RFC 6797)
- HTTP Public Key Pinning - HPKP (RFC 7469)
- Enforced Perfect Forward Secrecy mode
- Online Certificate Status Protocol - OCSP (RFC 4557)
- Scanning of Atom 1.0, RSS 2.0 and Sitemap files
- ICEcast/SHOUTcast streaming support

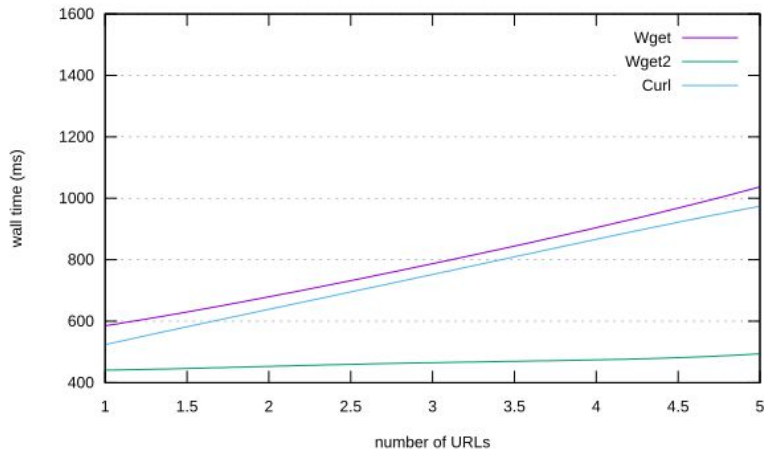
Improved Performance - HTTP/1.1

HTTPS with HTTP/1.1
Linux 4.12.8-2-ARCH x86_64 GNU/Linux, Intel(R) Core(TM) i7-5500U CPU@2.40GHz
ping RTT 112.632 to example.com
wget 1.19.1.79-c451e options: -q --no-config -O/dev/null
wget2 1.0.0 options: -q --no-config -O/dev/null --no-http2
curl 7.56.0-DEV options: -s -o/dev/null --cert-status --http1.1



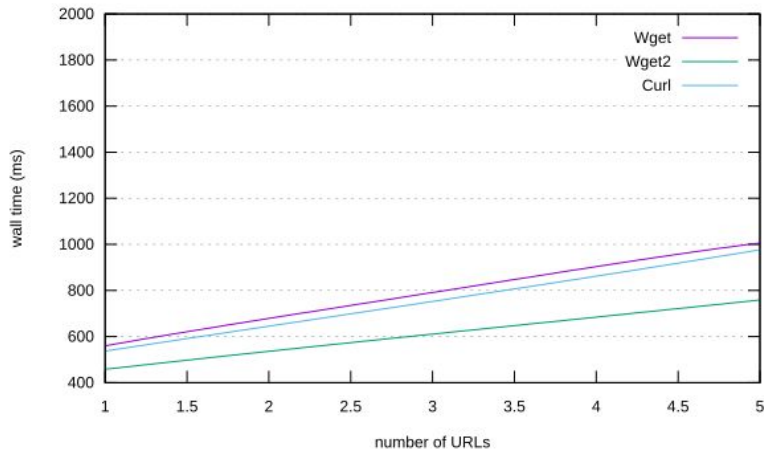
Improved Performance - HTTP/2

HTTPS with HTTP/2
Linux 4.12.8-2-ARCH x86_64 GNU/Linux, Intel(R) Core(TM) i7-5500U CPL2.40GHz
ping RTT 112.632 to example.com
wget 1.19.1.79-c451e options: -q --no-config -O/dev/null
wget2 1.0.0 options: -q --no-config -O/dev/null --http2
curl 7.56.0-DEV options: -s -o/dev/null --cert-status --http2



Improved Performance - HTTP/1.1 (1 Thread)

HTTPS with HTTP/1.1
Linux 4.12.8-2-ARCH x86_64 GNU/Linux, Intel(R) Core(TM) i7-5500U CPL2.40GHz
ping RTT 145.888 to example.com
wget 1.19.1.79-c451e options: -q --no-config -O/dev/null
wget2 1.0.0 options: -q --max-threads=1 --no-config -O/dev/null --no-http2
curl 7.56.0-DEV options: -s -o/dev/null --cert-status --http1.1





Future Enhancements

- GSoC: Plugin Framework (Akash Rawal)
- GSoC: Statistics Framework (Avinash Sonawane)
- GSoC: Test Framework using GNU Libmicrohttpd (Didik Setiawan)
- TLS 1.3 Support (Ander Juaristi)
- FTP / FTPS Support (maybe a separate wget2-ftp)
- WARC Web archive format (own project libwarc ?)
- Quic protocol (is there a GPL'ed libquic already ?)
- Certificate Transparency (CT)



Path to Initial Release

Wget2 works well as a daily driver and drop in replacement for Wget in most scripts

- Some bugs remain. Need more testers / bug reports
- Looking for contributors to reimplement existing switches from Wget 1.x
- Reviews / criticisms of Libwget API
- Documentation of Libwget API