

Exim and PCRE

How free software hijacked my life

Philip Hazel

University of Cambridge Computing Service

In case you don't know what Exim and PCRE are...

They are both free, open source software written by me

Exim is a Mail Transfer Agent for Unix-like systems (GPL)

- Drop-in replacement for Sendmail
- Runs on over 20 operating systems

PCRE is a C/C++ regular expression library (BSD licence)

- Supports regular expressions that are compatible with Perl
- Runs on Unix, Mac, Windows, ... (anything with C or C++)
- Used by open-source software: Exim, tin, Apache, PHP, ...
- Also in some commercial software: e.g. Mathematica

Why did I write Exim?

In 1995 we were running Smail 3 as our MTA

I liked it, but

- 7-year old code, not standard C, much hacked
- Unwanted legacy support for UUCP addressing
- Changing requirements
 - Virtual domains
 - More checking
- Difficulty of extending the code
- Decision: implement a new MTA

What were the original aims for Exim?

- To keep Smail 3's general approach
But lose the historical baggage
- To extent the capabilities
 - More controls (sender/receiver checking)
 - More reporting (logging, debugging features)
 - More generality (e.g. virtual domains)
 - More flexibility (e.g. in retry rules)
- Expect permanently connected hosts with good connections
- Assume “modern” OS and Standard C

How did Exim get released?

- Exim never had a formal “first release”
- Computer Science Dept begged an early copy
- They started running it, and told others
- We started running it on central servers
- I put the source on an FTP site
- Bug reports started arriving
- Requests for enhancement soon appeared
- Version went from 0.57 to 1.58 at one point
- Requests for a mailing list
- Planet Online (now Energis) put up a web site
- Mirrors volunteered themselves
- I started receiving T-shirts...

Why did I write PCRE?

Until 1997 Exim used imported code

- Original Henry Spencer library (1986)
- Simple regular expressions
- Perl's regular expressions were much more powerful
- But cannot disentangle Perl's code
- In the summer of 1997 I wrote PCRE
 - It provided Perl-style regex patterns for Exim
 - But I wrote it in C as a standalone library

What's the subsequent history of PCRE?

- I put it on an FTP site with a very liberal licence
People found it...
- Bug reports and requests for enhancements arrived
- Some upgrades to keep up with Perl
The biggest change was adding support for UTF-8
Not sure what will happen with Perl 6
- Now used by a surprising amount of software
- Development still ongoing
 - Unicode property support in Release 5 (September 2004)
 - Google contributed a C++ wrapper in Release 6 (June 2005)
 - Alternative matching function in Release 6 (June 2005)
 - New features from Perl 5.10 in Release 7.0 (December 2006)
 - Automake build system in Release 7.1 (April 2007)

What have been the main problems?

- A one-man band has plusses and minusses
For a long time I was the only person working on Exim code
- Documentation takes a lot of time, but is worth it
- The Exim FAQ should have been done earlier
(The FAQ has now evolved into a wiki)
- Less knowledgeable folk are now running MTAs
- The Exim books were a huge amount of work
- Debian's slow development cycle kept Exim 3 alive too long

Technical issues

- Variations in operating systems and shells
 - Signals – default restarting or not
 - Load average finding
 - Local interface finding
 - IP options
 - OS and architecture names
 - Argument types for library functions
 - Compiler pickyness
 - Things the C standard doesn't tell you
- Some OS keep changing rapidly (4 different IRIX configs)
- Testing an MTA is difficult (testing PCRE is easy)

Good things happened

- I have had strong support from my management and colleagues
- Exim and PCRE are an open-source contribution by
The University of Cambridge Computing Service
- The Exim community has been very supportive
Website and mailing list support (hardware and people time)
Mirrors for FTP and website
Many code contributions for Exim, some substantial
- Moving towards less dependence on me
CVS is in use for Exim development
Several regular comitters
- PCRE is in a SVN repository
But I'm the only comitter at present
- We now have a development system in Cambridge
It hosts the website, mailing lists, and CVS/SVN repositories

What should I have done differently?

- There have already been several rethinks
 - Memory handling
 - Lookup handling
 - Negative matching options
 - Merging of directors and routers (no special domains)
 - Use of ACLs for input control
- Use less C I/O
- More radical changes to the Smail 3 conception
- Think bigger and more generally

Would I do it again?