OPENING THE SOURCE REPOSITORY WITH ANONYMOUS CVS

Charles D. Cranor

AT&T Labs-Research

chuck@research.att.com

Theo de Raadt

The OpenBSD Project

deraadt@openbsd.org

OUTLINE

- **Introduction and motivation**
- **Background**
- **Anonymous CVS: design and implementation**
- **Other open source repository tools**
- **Conclusions**

Open source: making inroads

- **projects: Linux, BSD, GNU tools, Apache, Mozilla,...**
- **key attributes:**
  - source code freely available
  - open license
- **advantages: promote reliability/quality via:**
  - independent code review
  - rapid evolution

Only a relatively few users take advantage of having access to source

- few people download it, fewer read it
- pre-compiled distributions quite successful
- open source developers do access it

Standard "Open Source" only partly meets the needs of open source developers

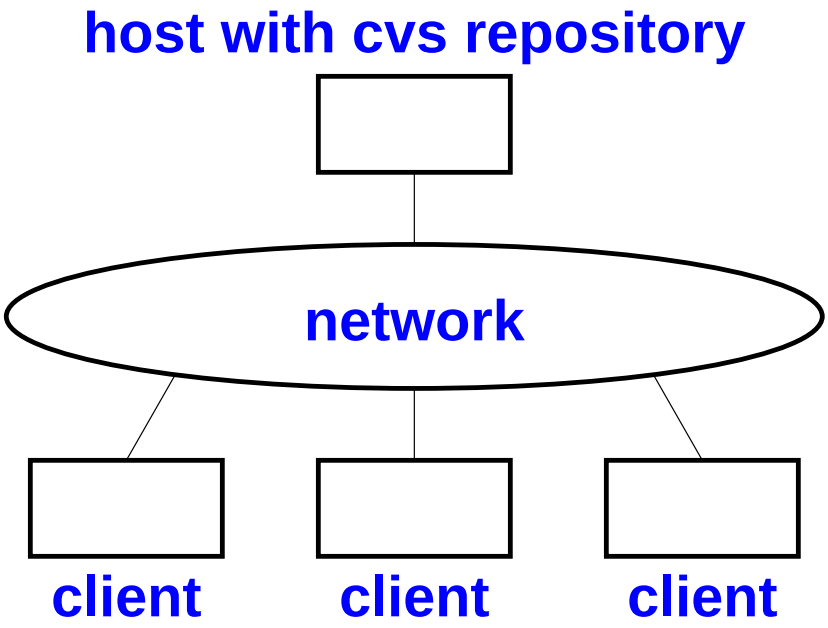
- **missing features**

  - access to old versions of code
  - annotated per-file modification history
  - set files to a distribution or date
  - get current snapshot
  - merge in local changes

- **features provided by source control system**

  - local: SCCS, RCS
  - networked: CVS

CVS was not designed to be open (1995)



- **Usage requirements:**
  - account on repository host
  - write access to repository
- **Only select group of privileged developers can access CVS repository**
- **Counter to open source philosophy**

Fall 1995: OpenBSD project started

- **goal: open access to CVS repository**
  - attracts users
  - makes it easier to download, debug, and manage source tree
  - easier to learn about the evolution of code

- **result: Anonymous CVS service**
  - Internet users have read-only access to data in repository...
  - extends "Open Source" concept to "Open Source Repository"

Traditional source distribution

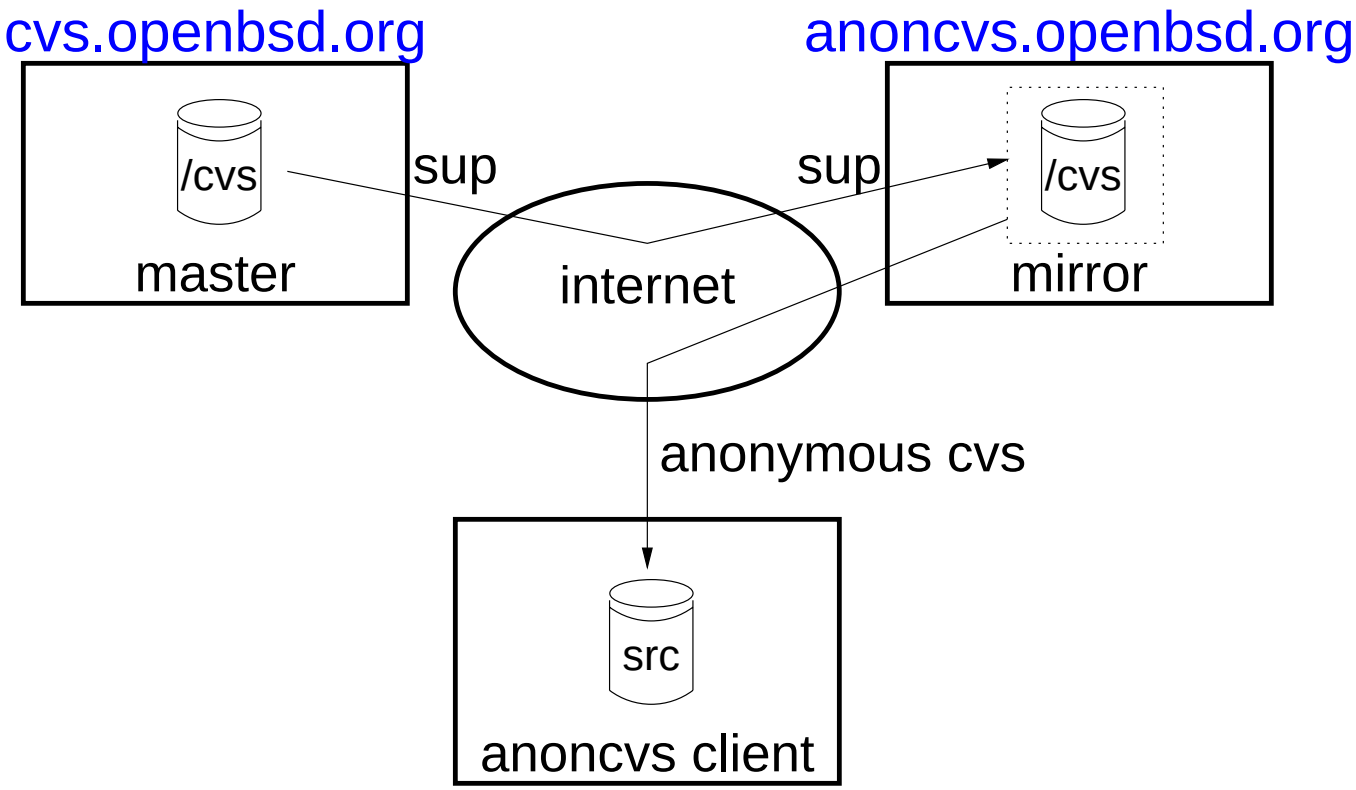
- **USENET comp.sources.***
- **Anonymous FTP / web**
- **sup**
- **rsync**
- **CTM**

All: Open source, but not open source repository

Design goals:

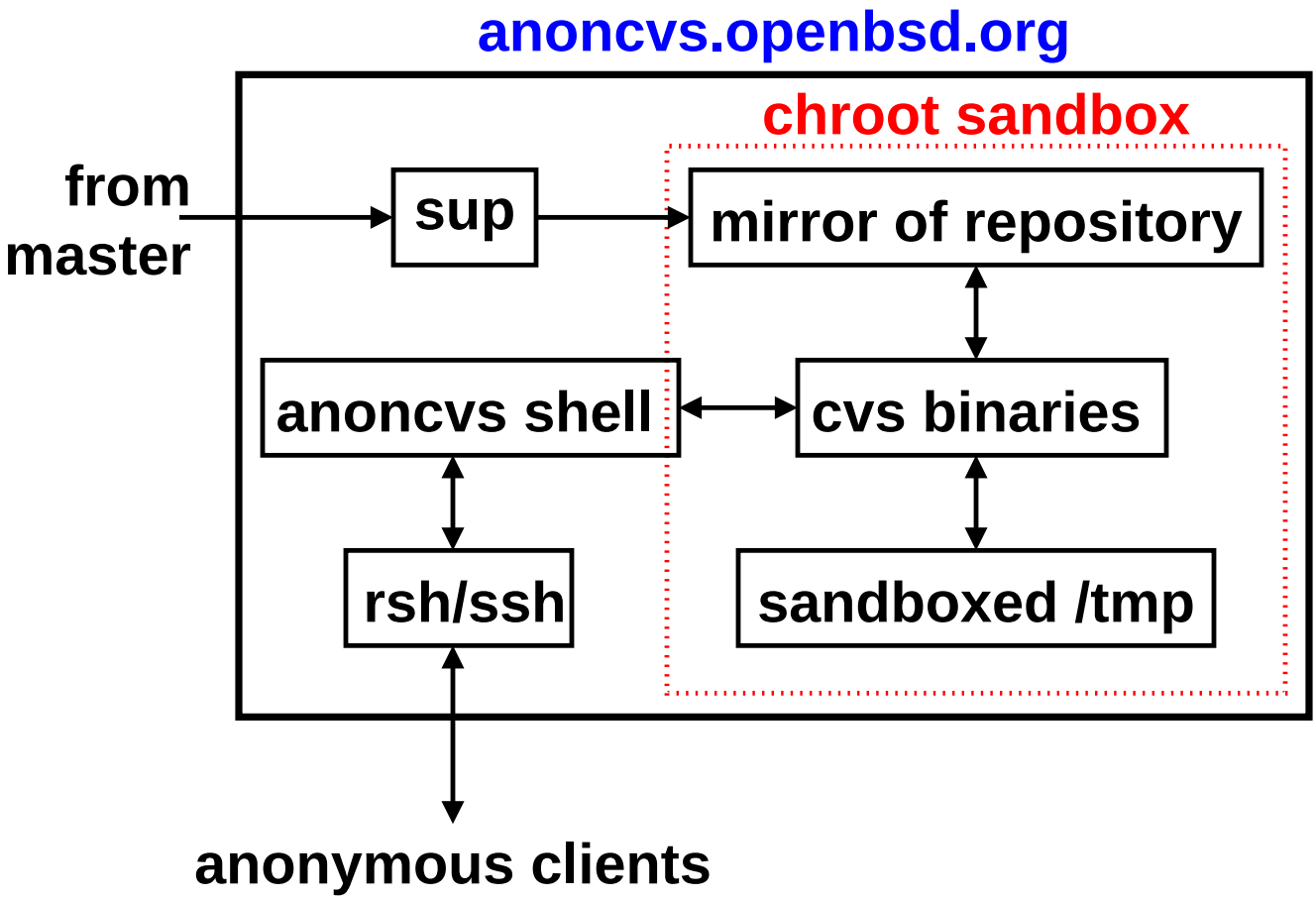
- **security**
- **efficiency**
- **convenience**

ANONYMOUS CVS DESIGN



- **Put AnonCVS on secondary server machine**
  - replicate repository on AnonCVS server
  - control anonymous load on main server
  - no direct anonymous access to master repository

ANONYMOUS CVS DESIGN



- cron gets repository (via sup or rsync)
- mirror owned by non-priv account
- "anoncvs" account: no password, captive shell
- cvs runs in chroot sandbox environment

Implementation issues

- **cvs required writable log file**

  - added CVSREADONLYFS environment variable

- **file locking**

  - CVS locking not an issue with read-only repository
  - incomplete CVS file: cannot happen

    1. create temporary file
    2. write complete file data to tmp file
    3. rename() tmp file to real file [atomic]

  - SUP removes a CVS file (should not happen)
  - old/new mix: possible (even with standard CVS)

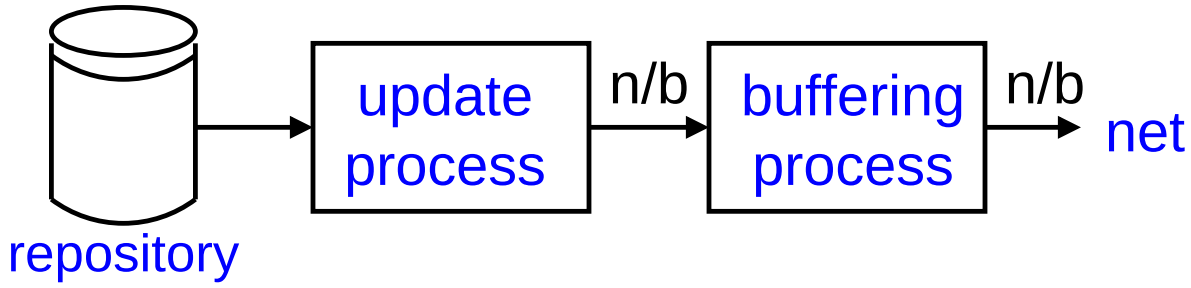
Implementation issues (cont.)

- **network flow control problem**

  - CVS design goal: minimize locking time
  - problem:

    1. lock CVS files
    2. send update to remote system
    3. unlock CVS files

  what if we block in step 2? (network flow control)



problem: no limit on buffering process' buffer size

solution: limit buffer size, ignore locking

Anonymous CVS deployment

- **Fall 1995: first anonymous CVS server (wustl.edu)**
- **Currently OpenBSD has 20 AnonCVS servers**
- **Usage: 2000 transactions/week (main server)**
- **Attracted contributors**

After Anonymous CVS

New Open Source Repository Tools created

- **CVS' pserver (cvs developers)**
  - adds anonymous support to CVS
  - uses special CVS server ports
  - user interface requires login/password
  - often does not run in chroot() environment
  - now included with CVS

- **CVSWeb (Bill Fenner, FreeBSD)**
  - browse CVS repository via web client
  - no local CVS tools required
  - graphic user interface to CVS

After Anonymous CVS (cont.)

- **CVSSup (John Polstra)**
  - current state of the art in Open Source Repository tools
  - can distribute repository or source tree
  - uses highly efficient streaming protocol
  - knows file formats:
    1. CVS/RCS files
    2. log files
    3. unknown (uses rsync algorithm)
  - can merge into local repository
  - has graphic user interface
  - requires Modula3 to compile

CONTRIBUTIONS

- **we have extended "Open Source" to the next level**

  Open Source => Open Source Repository

- **positive effect of Anonymous CVS (e.g. OpenBSD)**

- **Anonymous CVS helped lead to the introduction of new Open Source Repository tools**

- **Many large projects have embraced Anonymous CVS**

  - Ecgs, FreeBSD, Mozilla, Apache, etc.