

Open Source Internet Filtering



John Brice
Meadville Public Library

Meadville Public Library
meadvillelibrary.org

What is Open Source?

- Free as in freely modifiable
- Free as in free of charge
- Free as in free of licensing hassles
- More info: see www.opensource.org

Open Source Filtering Applications

- SquidGuard
www.squidguard.org
- DansGuardian
www.dansguardian.org

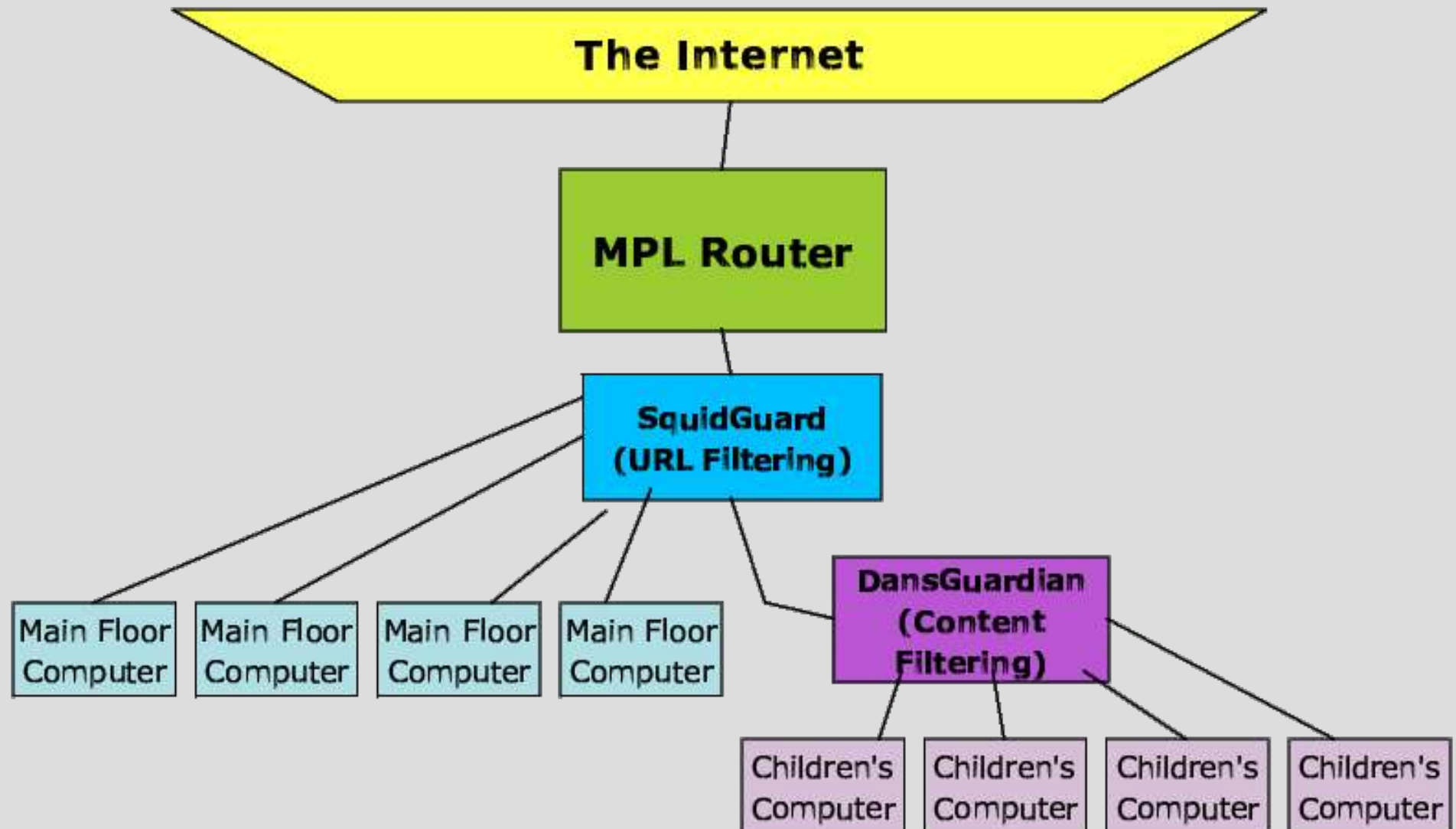
SquidGuard

- Filters by URL (or contents of URL) only through the use of:
 - blacklists
 - exception lists
 - and keyword lists

DansGuardian

- Filters by website content through the use of keyword/phrase lists, blacklists, and exception lists
- Can filter by file/mime type (to prevent downloads)
- Can also use PICS content ratings system

MPL's Filter Architecture



Advantages of These Filters

- Software is free
- Library staff has full access to blacklists, exception lists, and expression lists
- Library staff can customize all lists
- Can be monitored & maintained remotely
- Server-based; makes network-wide updating easy

Advantages, Continued

- Filter behavior is customizable & controllable
- Can be run on many free operating systems: Linux, OpenBSD, FreeBSD, etc.
- Client independent; can be used through any browser
- Doesn't require an expensive server; scalable
- Rock solid (reliable & stable)

Advantages, Continued

- Support through mailing lists is usually excellent, and often directly from the author of the program
- Commercial support is available
- Can be used on as many computers as you need to without paying per-seat licensing fees
- Graphical maintenance tools are available

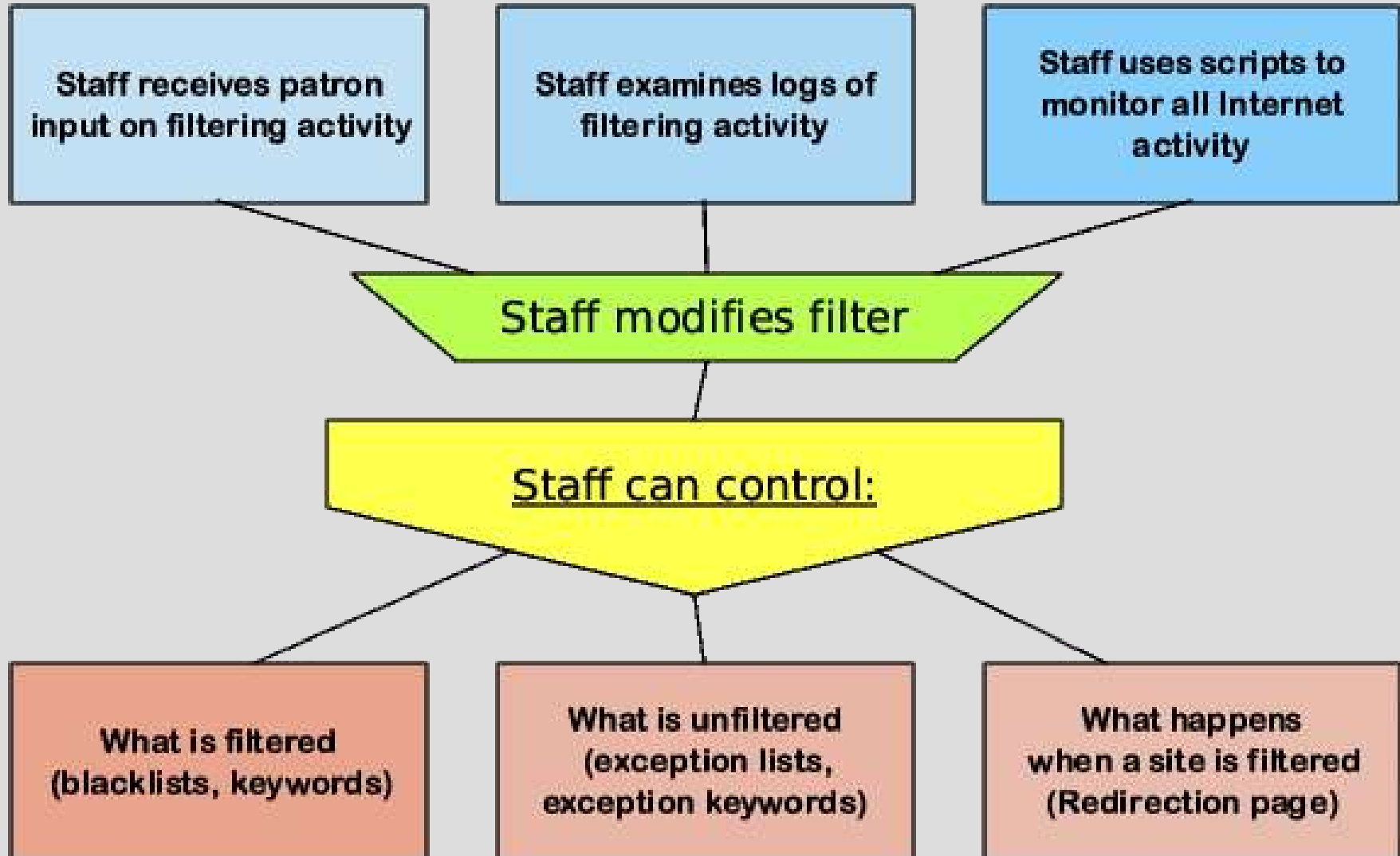
Disadvantages of OS Filters

- Steep learning curve
- Requires more knowledge than most commercial filters to install and maintain
- More time intensive to install & operate

Compared to Commercial Filters

- Commercial Filters:
 - Are not free
 - Won't allow you to view blacklists directly
 - Don't always work as expected
 - Won't always allow staff to control what happens when a site is filtered
 - Are often client-based

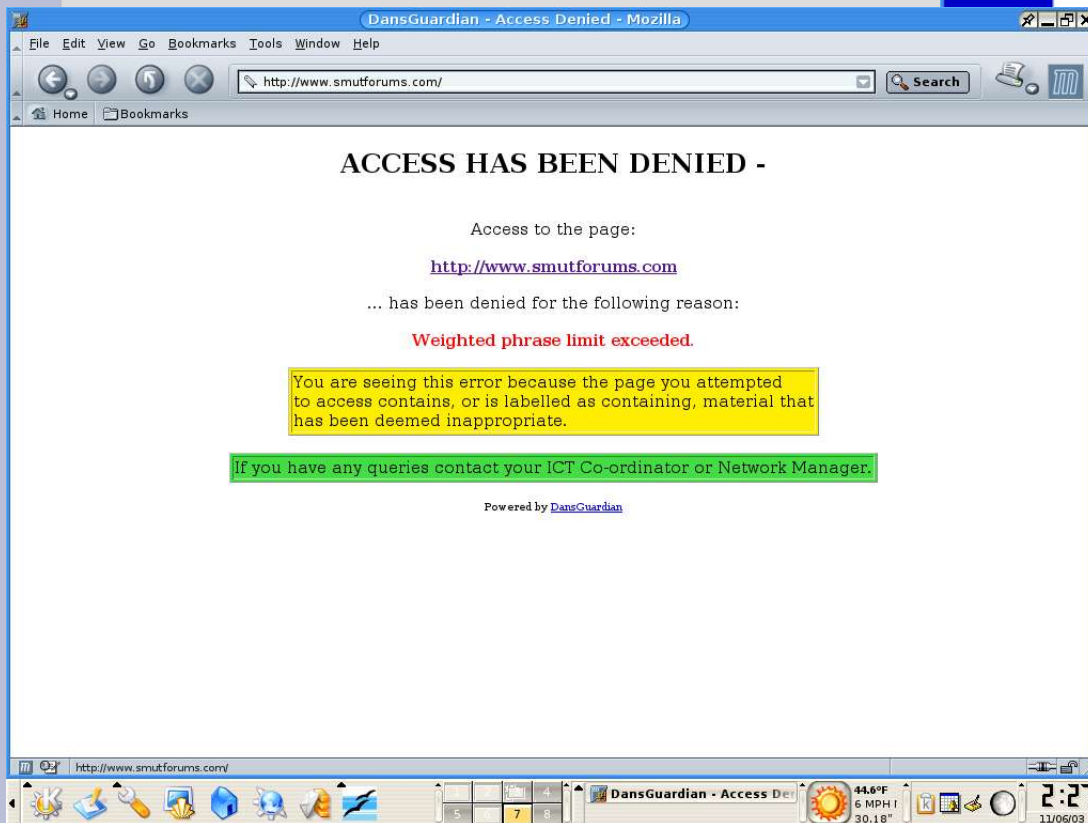
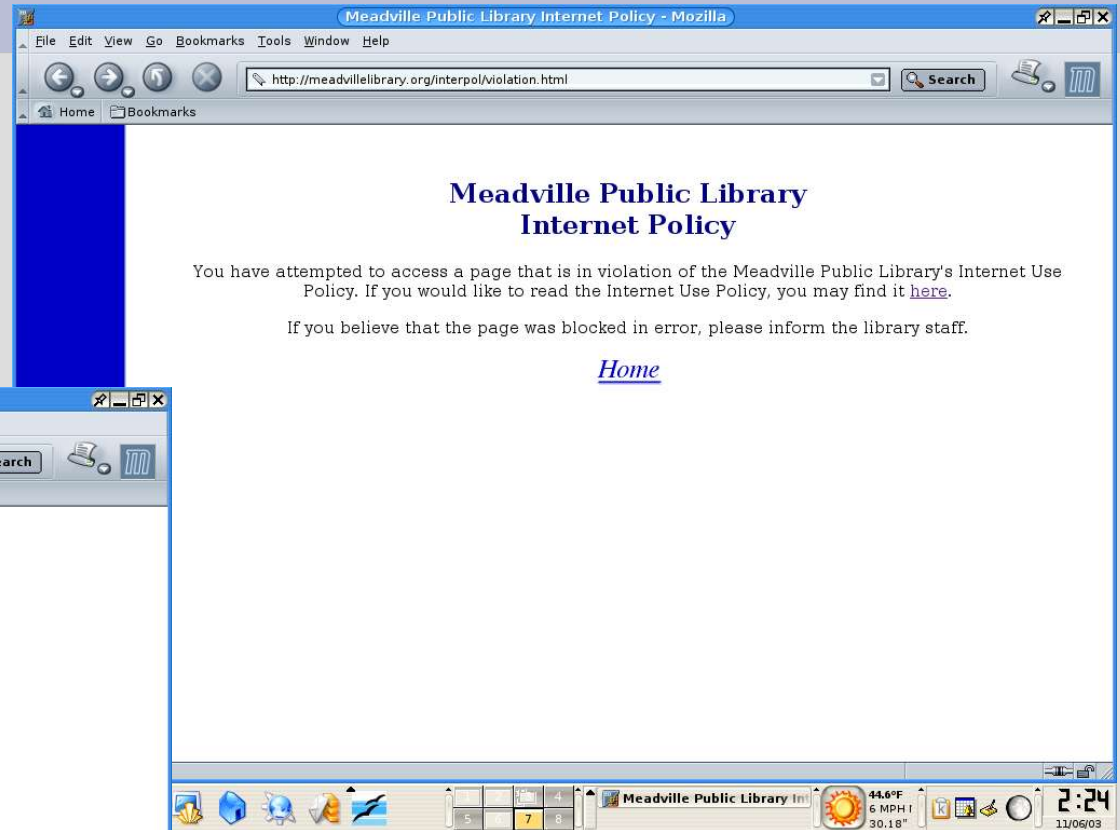
Filter Maintenance



Disabling the Filter

- The filter can easily be disabled by library staff by changing the browser settings
 - if properly secured, the browser will revert to its filtered state once closed and reopened
- Individual websites can easily be unfiltered by adding them to the filters' exception lists

Custom Redirection Pages



Where These Filters Could Be Used:

- Libraries with at least one staff member with the knowledge and time to install and maintain the filter
- Libraries with the above, or with someone outside the library who can install and maintain it (a District/System tech person, a consultant, etc.)
- A knowledgeable person should be able to install & configure the OS & software in about 8 hours

Hardware Scalability

- MPL's server:
 - 1.2 GHz processor
 - 512 MB RAM
 - 10 to 15 users
 - 90 to 100 % idle
- At smaller libraries we use \$199 Walmart computers (900 MHz, 128 MB RAM)

Study of Our Filter's Accuracy

- Used a random sample of 600 sites from our blacklists (out of 345,155)
- Sites evaluated by 3 staff members

Study Results

- Prurient: 49%
- Non-prurient: 5%
- Nonfunctional: 33%
- Possibly Prurient: 1%
- Domain Park: 4%
- No Web Site: 5%
- Utility or Configuration Page: 0.05%
- Error: 2%
- Hosting Site: 1%

Analysis of Findings

- If no longer valid sites are eliminated, total overall effectiveness is 83%
- Many of non-prurient sites used prurient-style names, either available for resale or featured directories of prurient sites
- Overall more accurate than commercial filters

For Further Information

- www.meadvillelibrary.org/os features:
 - More presentations on Open Source
 - Links to articles on Open Source
 - Information on how to install & maintain SquidGuard and DansGuardian
 - Links to other related websites