

Extending Tor Network Metrics

Kevin Berry

Abstract

I intend to extend the Tor metrics portal (specifically Ernie) to store long-term network statistics, as well as simplify the process of searching large data sets from the relay descriptors.

1. The Project

The basis for this project is idea #2 on the list - "Help track Tor overall network status." This project will be concerned with generating and analyzing long-term Tor network status and statistics. While there are visualizations and metrics in place to view recent network activity, there is no such thing to view even higher-level network activity such as node attrition and new node rates.

The Ernie project powers the metrics portal to scrape data from various sources, generate graphs, and serve statistics. It currently serves static html from which graphs are generated periodically. The idea is to move it over to a web application where graphs and metrics can be generated on demand. In addition, moving Ernie to a dynamic setup will expedite the process of publishing directory data. While the relay descriptor data sets are extremely large, a data warehouse concerning a few statistics will be put in place. Furthermore, and perhaps most importantly, a search feature will be implemented regarding the data sets. This may be useful for debugging, and finding specific information about a node at a given time.

A few tools I have in mind as a starting point:

- Apache Tomcat - JSP & Javadoc - to work with existing java metrics code
 - Build upon Spring MVC framework?
- PostgreSQL - The database already in place
- Apache Ant - For building/deployment
- R - For graphing and working with existing code
- Eclipse, vim
-

A loose schedule:

- Week 1-2 - Research, research, research. Begin some minor work
- Week 2-4 - Port Ernie and current functionality to Tomcat framework
- Week 5-8 - Work on analyzing data sets for statistics -> Java code
- Week 8-9 - Work on search features regarding work and statistics from previous weeks
- Week 9-10 - Automation of data set publishing
- Weeks 11-12 - Overflow period, fixing bugs, maintenance.
-

2. Code Samples

Please view my website at <http://kevinjberry.com> for a few project code samples. Unfortunately my largest project (flu clinic software) is proprietary. I may post some code, though.

- BranchAndBound - A java implementation of a branch and bounding type algorithm to find the lowest combination in a matrix.
- Sprite Cutter - A C++ tool to automatically cut a single-image sprite sheet into discrete images, as well as publish CSS and XML sprite vectors. Is currently being made into a web application. Works with the OpenCV and CvBlob libraries. My "scratch an itch" project. It's still a work in progress.
- Distributed Computing from the Browser - PHP, JavaScript, and SQL project that collaboratively reverse hashes a hidden list of words with the help of people browsing the internet.

A few others I've done for classes:

- HTTP proxy server in C
- Bash-like shell with backgrounding and jobs functionality in C
- SMTP and ESMTP application in C
- A scheme-like language parser and semantic analyzer in Java

3. Why I want to work for Tor

Open source software has given a lot to me, and I have been considering ways to give back. I use and depend on free software every day. Perhaps more importantly, I've learned the most from using and interacting/tinkering with free software. Without getting overly philosophical on the merits of absolutely, 100%, amazingly *free* software, I'd like to thank the people that make it a reality, as well of the likes of evangelists like Richard Stallman and John Perry Barlow. Free software rules.

Tor is one of the strongest representations of anonymity on the internet. Anonymity is one aspect that makes cyberspace very powerful (and sometimes scary?) - one need not look further than the Iranian elections, in which Tor in particular received a lot of press. Tor is credited as a liberalizing technology which upholds important ideals. And, I want to learn.

4. Software development team experience

I've worked in a small team environment for a medium sized project (~18k lines of code and html) involving flu clinic software. However, it was more of me being mentored after taking over the project. The goal was to finish in 8 weeks. Working by oneself can be difficult, but everything was finished by the deadline and the clients were very pleased. They used to manage about 20 flu clinics before the software. The year after they used it, they were able to successfully run over 400 flu clinics over a larger geographical area.

Working with knowledgeable people is invaluable. From my exchanges so far with Karsten I can tell he is both enthusiastic and knowledgeable. Fortunately I also have lots of smart classmates. Group projects with them have always turned out well.

I'm frequently asked for help regarding GNU/Linux. I guess I'm the class Linux geek. I've so far used the more basic features of svn and git for version control to work on my team projects.

5. Availability

I will dedicate myself full-time with Tor. I have a good set-up at home and may periodically work from my father's office to have a better working environment.

My schedule so far would look like:

- April 26 - My application gets accepted!
- May 7 - School ends - I relax for a bit then start researching/working
- May 24 - GSoC officially starts
- August 20 - I finish project. Continue working with Tor. I go back to school 3 days later.

6. After summer?

Maintenance is one of the most important aspects of a software project. It's often cited as the highest cost. I intend to maintain and possibly improve anything I work on. Certainly, bug fixes regarding any software I produce will be my responsibility. But, one of the biggest reasons I would like to continue working with Tor is because I would love to see my software being used by people.

7. Communication

I intend to be available on IRC, Aim, and via e-mail. I plan to stay in close touch with my mentor and others at Tor. This is because without close communication, the project can probably go "off the rails". Making something really useful also means sticking to specifications and listening to the higher-ups. Requirements and minor details are bound to change, and it is important to keep everyone in the loop as far as progress and changes are concerned.

So far, interaction with Tor has been good. My exchanges with Karsten have been pleasant, and he has welcomed and answered any questions I have had, as well as given me plenty of advice!

I also intend to blog periodically about my progress on my personal website.

8. Education and Focus

B.S. Computer Science, Villanova University May 2011. I'm currently a junior. My last term project focused on distributed computing via the web browser.

9. Contact

IRC (freenode): kjbbb (usually idling)

E-mail: kevin.berry@villanova.edu, xckjb88@gmail.com

Aim: xckjb

10. Anything else?

I may involuntarily grow a Unix beard.