Measuring the Tor Network
— Evaluation of Relays from Public Directory Data —

Karsten Loesing
karsten@torproject.org

Tor Tech Report 2009-06-001*
June 22, 2009

Abstract

This document contains the results of an evaluation of the directory archives containing network status consensuses, router descriptors, and extra-info documents of relays from February 2006 to June 2009.

*This report is superseded by: Karsten Loesing. Measuring the Tor Network from Public Directory Information. 2nd Hot Topics in Privacy Enhancing Technologies (HotPETs 2009), Seattle, WA, USA, August 2009.

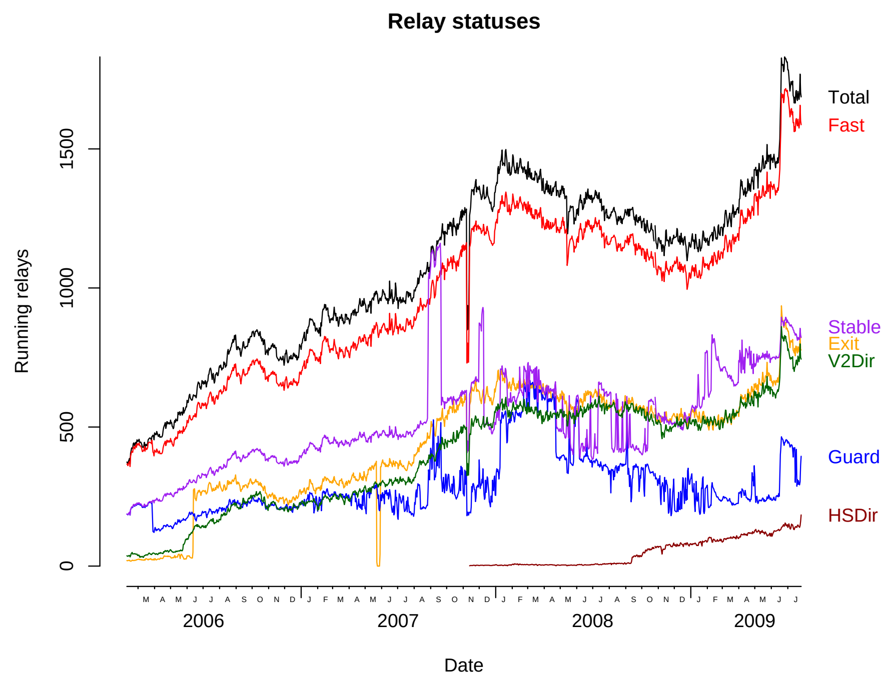

Figure 1: The lines represent the number of *running* relays with certain flags as listed in network status consensuses from February 2006 to June 2009. This diagram can be useful to decide whether conditions to assign certain flags might require modification. It is questionable whether the Fast flag is assigned too often to make a good distinction between fast and slow nodes. The volatility of the Stable and Guard flags might be the result of problems in the directory voting process. The average number of guard nodes has become rather low, given that these nodes need to push one third of all user traffic; changes to the requirements for becoming a guard node should be considered.

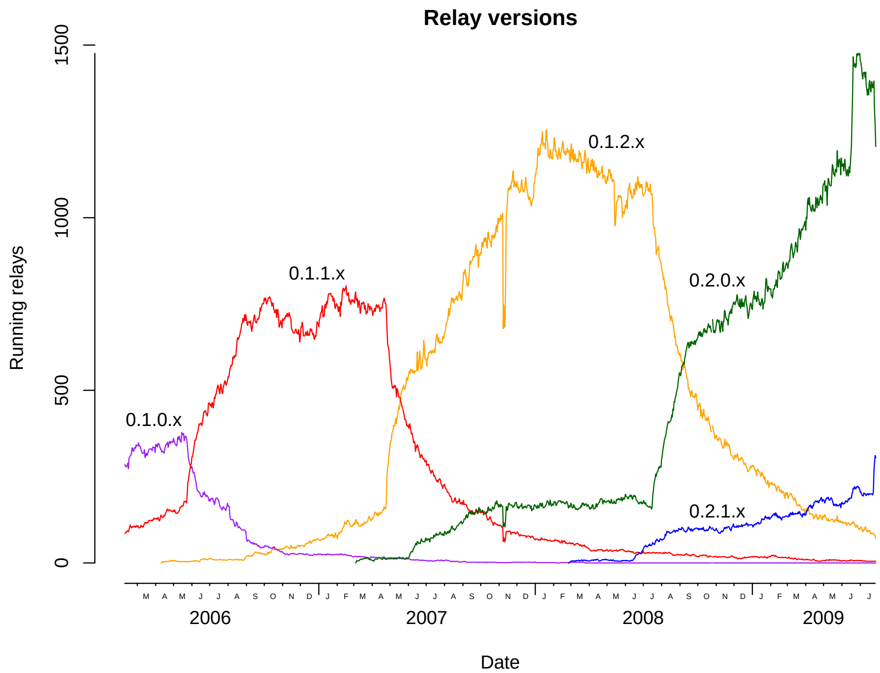

Figure 2: This graph visualizes the version life cycle of Tor relays as declared in router descriptors. Each version life cycle can be subdivided into an alpha and release candidate phase (April 2006 to April 2007 for 0.1.2.x), a stable phase (April 2007 to July 2008), and an out-of-date phase (July 2008 until today). For all versions there is an upper limit of approximately 200 relay operators running alpha or release candidate versions. There is no visible increase when versions are moved from alpha state to release candidate state (March 2, 2007 for 0.1.2.x, February 24, 2008 for 0.2.0.x). The stable phases for all versions show that it can take months until relay operators switch from an out-of-date version to the current stable version (April 2007 to end of 2007 for 0.2.0.x). Accordingly, the out-of-date phases show that old versions are used even years after new stable versions are available (0.1.1.x still in use in 2009). These results indicate that a semi-automatic updating mechanism is desirable.

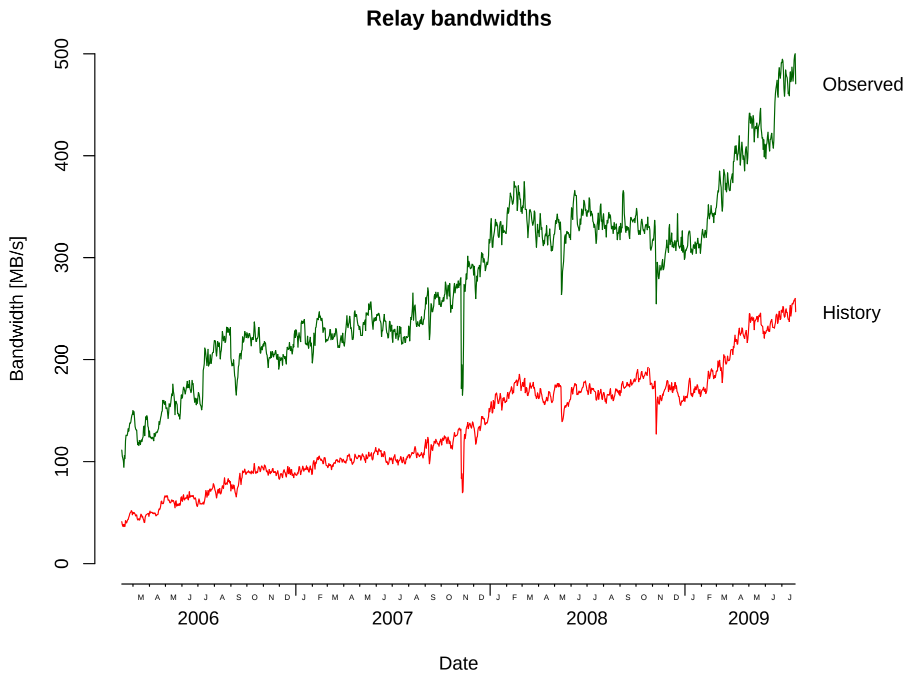

Figure 3: Relays report their observed bandwidth capacity and bandwidth usage to the directories. The bandwidth capacity (green line) is the maximum bandwidth as observed over any ten seconds in the past day. Bandwidth usage (red line) is calculated as the total size of relayed bytes in 15-minutes intervals over the past day. The graph shows that roughly half of the available bandwidth capacity is used by clients—and that the other half remains unused. Further investigations should focus on relays which leave most of their bandwidth capacity unused. Either the approach to measure capacity by 10-seconds bandwidth peaks is infeasible, or clients do not make us of the available bandwidth sufficiently.

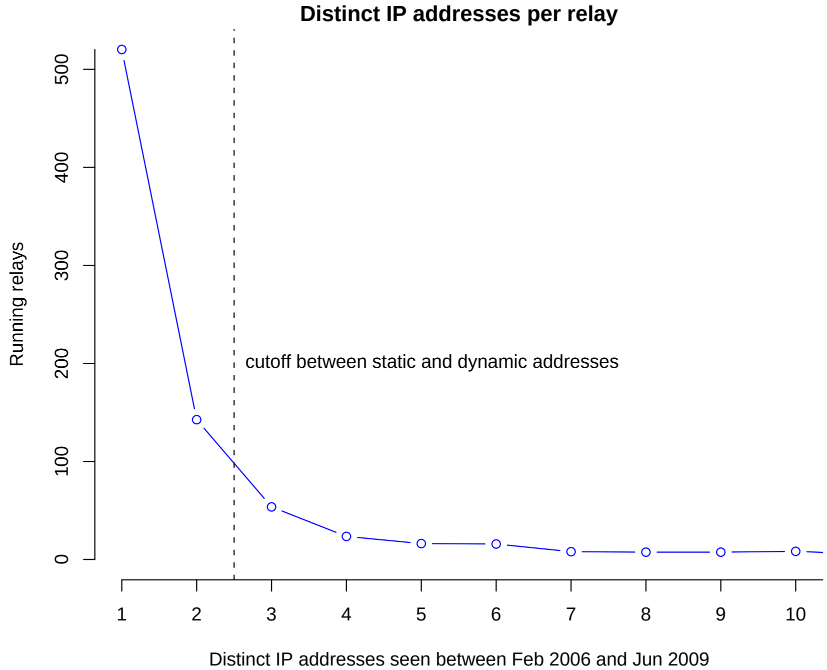

Figure 4: Some relays run on static IP addresses, some on dynamic addresses. This graph visualizes how many relays in a typical consensus have been seen with how many distinct IP addresses between February 2006 and February 2009. For this analysis, a somewhat arbitrary cutoff between static and dynamic addresses is made between 2 and 3 IP addresses: Relays seen with at most 2 addresses are considered to run on static IP addresses, the others on dynamic IP addresses.

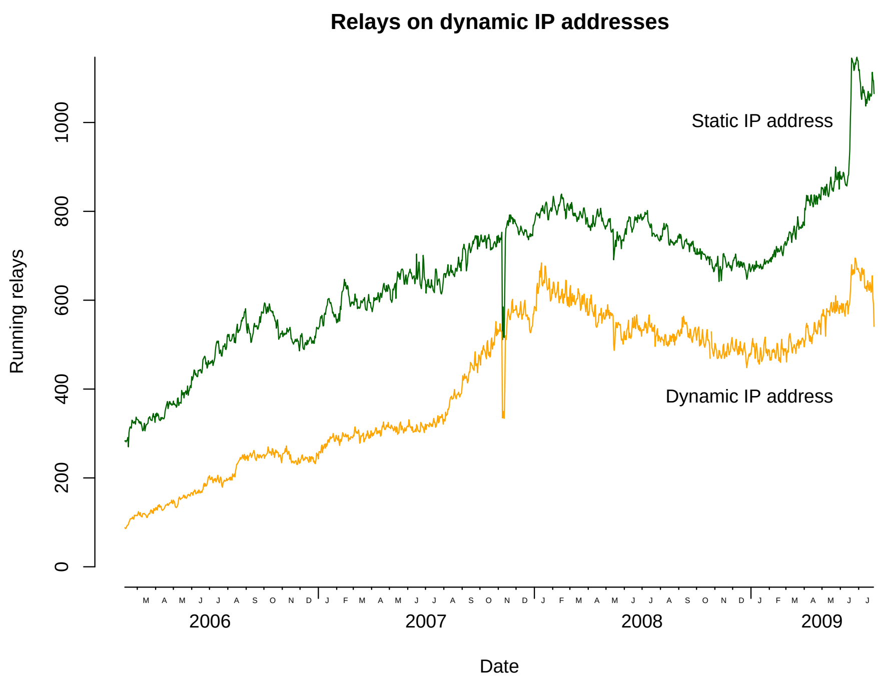

Figure 5: This graph shows the number of relays running on static (green line) or dynamic IP addresses (red line). The high number of relays on dynamic addresses indicates that efforts should be taken to make new relay addresses available to clients more quickly. Otherwise, a certain share of relays is unreachable for clients, leading to under-utilization of available bandwidth.

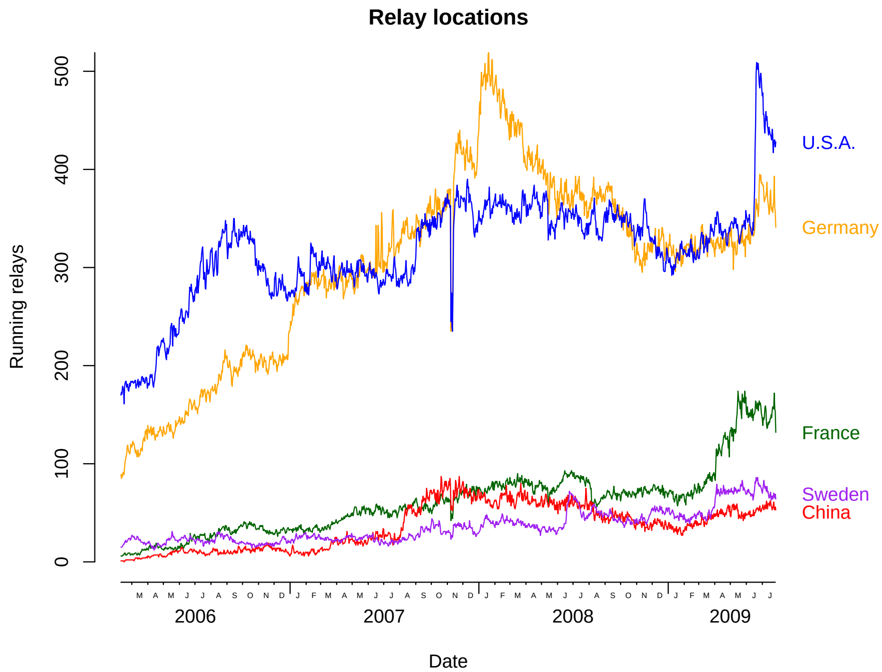

Figure 6: This graph shows the numbers of running relays in the top-5 contributing countries. The most visible trend is that the number of German relays suddenly stops growing in January 2008 and significantly shrinks over 2008. This trend might be the result of data retention laws. Another interesting trend is the sudden increase of U.S. relays in June 2009, possibly a result of the Iranian elections.

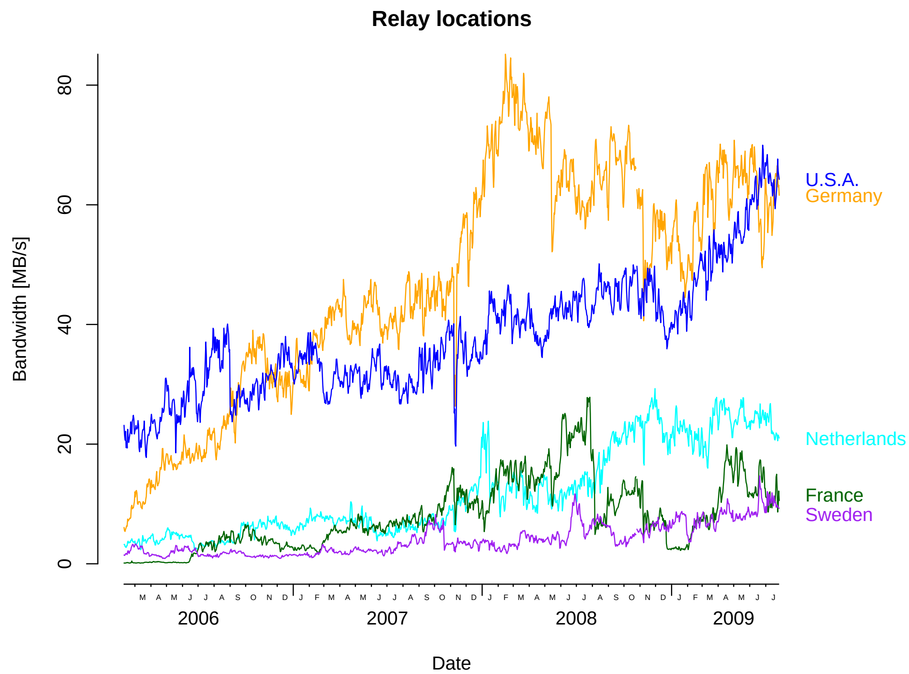

Figure 7: This graph shows the top-5 contributing countries, this time by bandwidth usage. The pattern for German nodes in 2008 is similar to the figure above. Another pattern is that French relays have suddenly seen less usage in July 2008. Finally, the Netherlands are the third largest provider of bandwidth, even though it did not show up in absolute numbers in Figure 6.