Data Analysis of Tor Tickets

Marie Yau

March 18, 2018

1 Introduction

This is an analysis of the Tor bug tracking system, based on Trac data extracted for the month of Jan 2018. The data exported did not have a critical metric: ticket lifespan (ticket creation time - ticket resolution time). This information was manually extracted from Trac GUI, calculated for each of the 320 entries, then imported into R for analysis. Results indicate the majority of bugs are in the Tor Browser or Tor. A surprising discovery is the large variation in ticket lifespans for each ticket owner. Further analysis with additional data is needed to determine the cause of this variation. The result of this analysis could lead to a faster, more efficient bug tracking system.

2 Analysis of Tickets

2.1 Status of Tickets

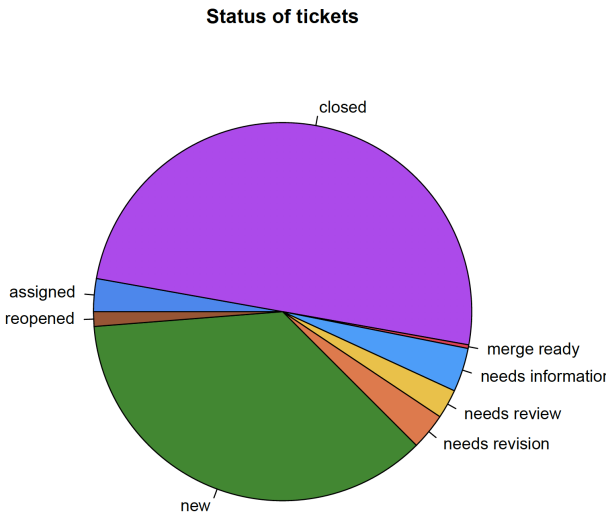

Figure 1: Pie chart of ticket status

As we can see, most of the tickets from January, 50%, were closed. However, 36% of tickets remained opened as of March 12.

2.2 Ticket Components

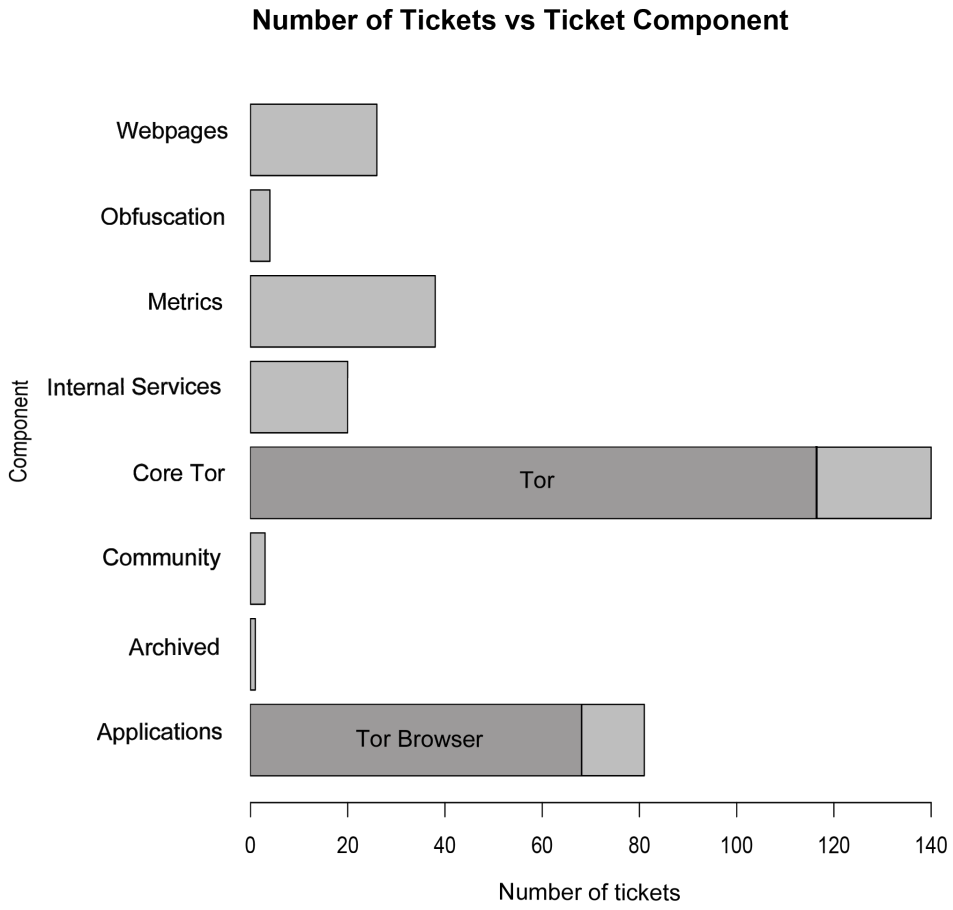

Figure 2: Bar chart of ticket components and sub-components versus number of tickets

The majority (69%) of the ticket reports problem with either Core Tor or Application. Within Application, the 83% of issues are Tor Browser (sub-component) related, and 81% of issues with Core Tor involved problem with Tor (sub-component).

2.3 Ticket Severity

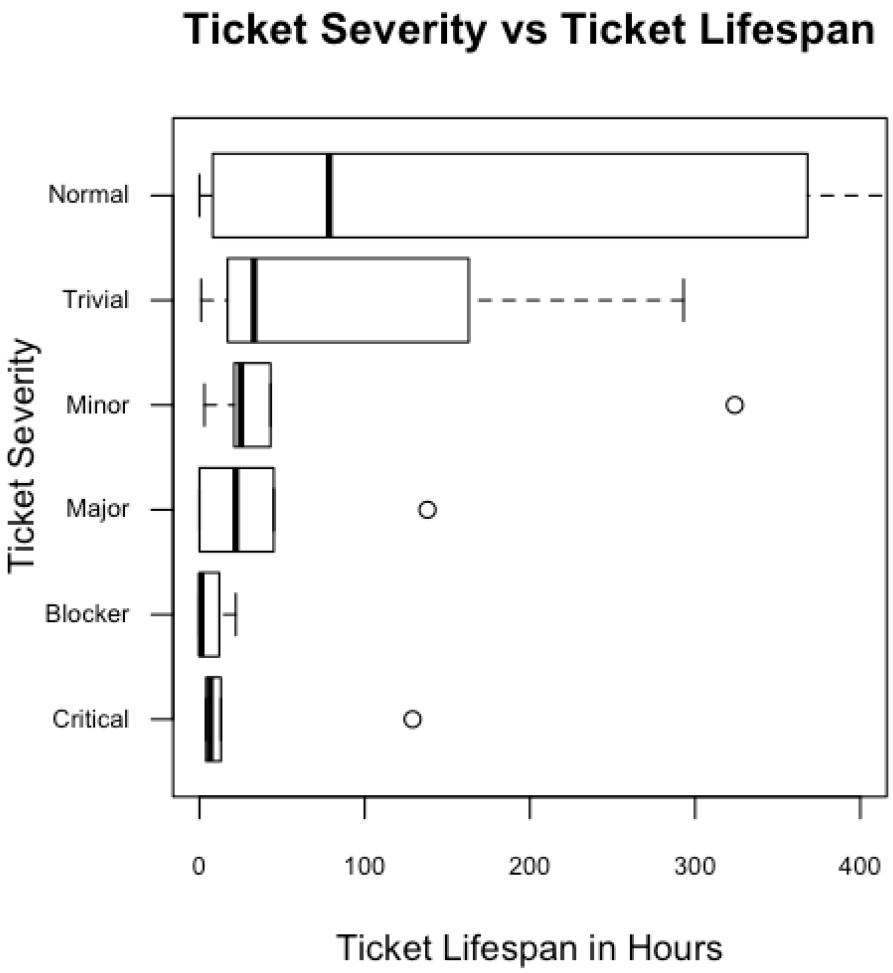

Figure 3: Box plot of tickets severity versus tickets lifespan

As shown in the above boxplot, ticket severity does impact the resolution time. We can confirm that tickets marked as Critical has the shortest resolution while Normal tickets have the longest resolution time.

2.4 Ticket Priority

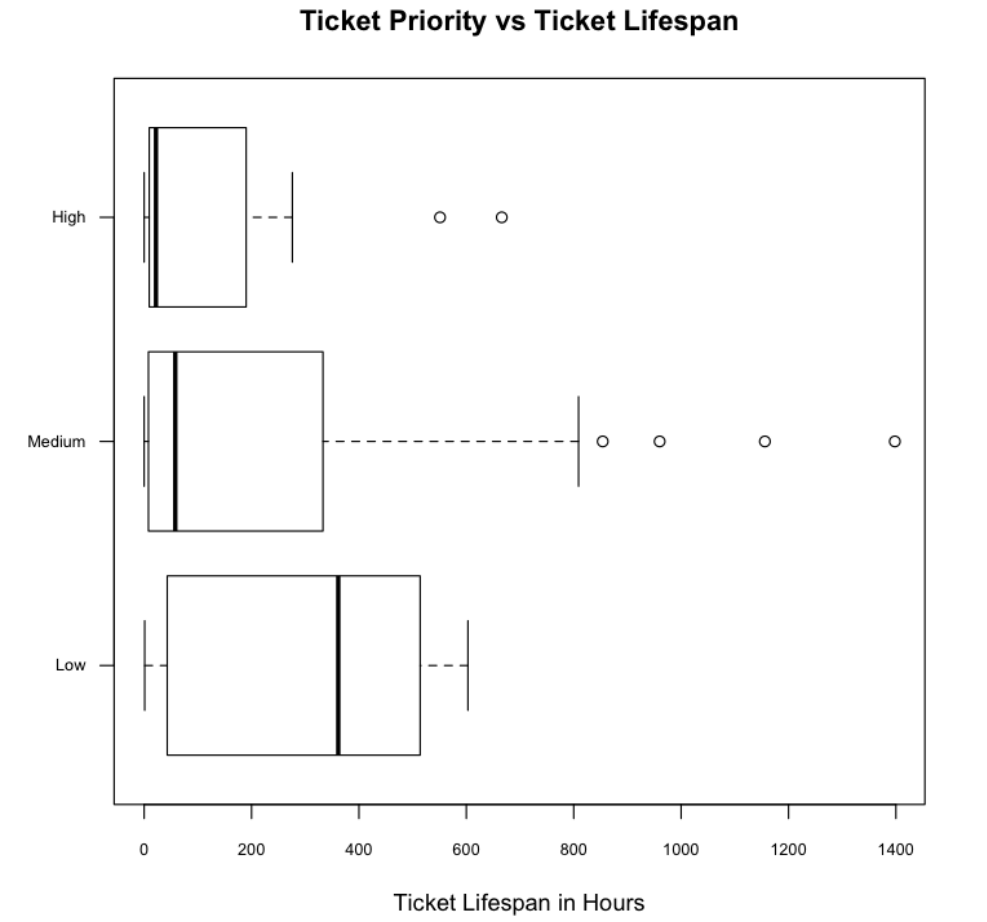

Figure 4: Box plot of ticket priority versus ticket lifespan

Ticket priority also influences the resolution time. With high priority tickets resolving quickest and low priority tickets do have the longest resolution time.

2.5 Ticket Lifespan

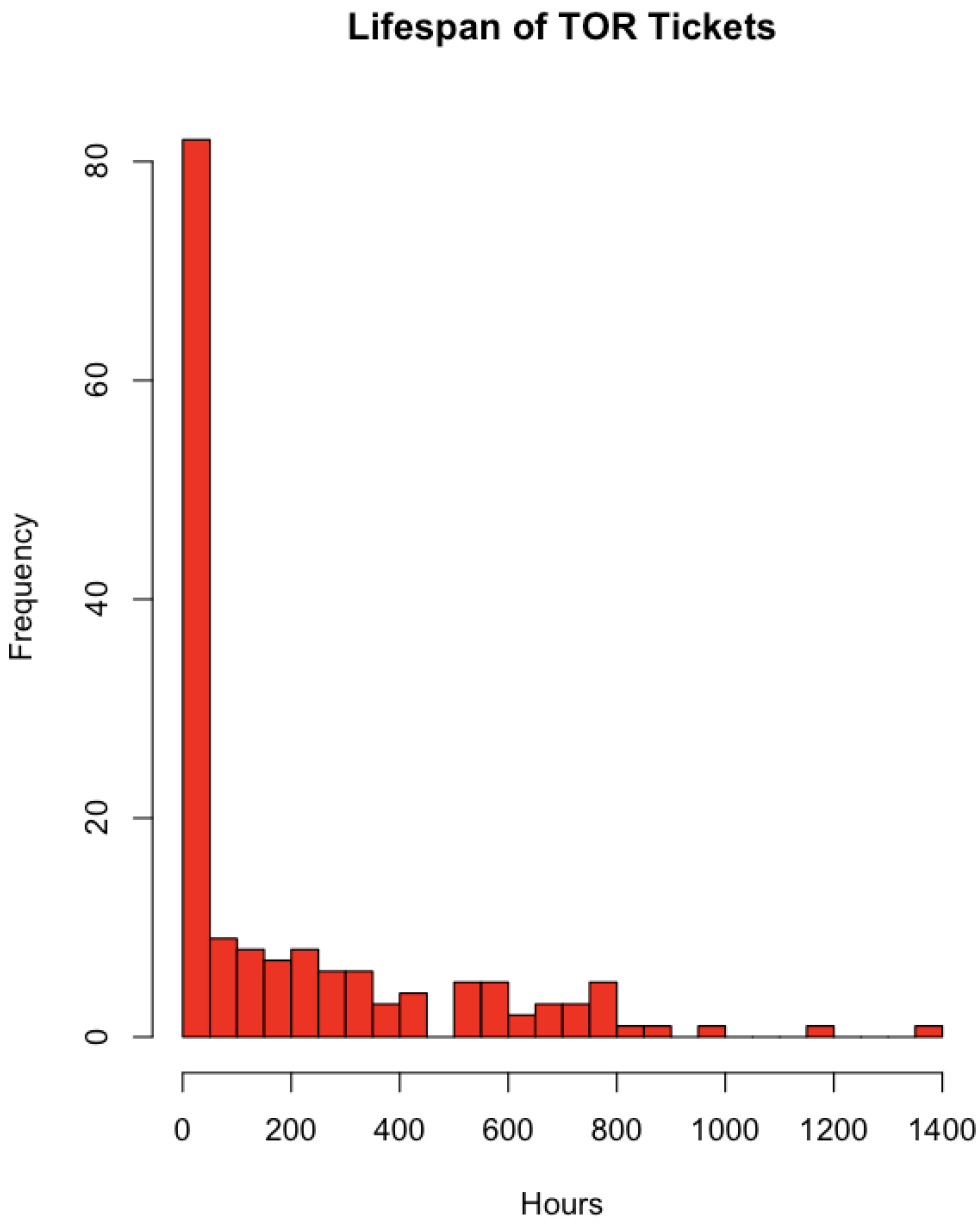

Figure 5: Histogram of ticket lifespan in hours

The above histogram displays the overall lifespan (from creation to resolution time) of all closed tickets. The majority of the ticket gets resolve within two days.

2.6 Day of the week tickets open

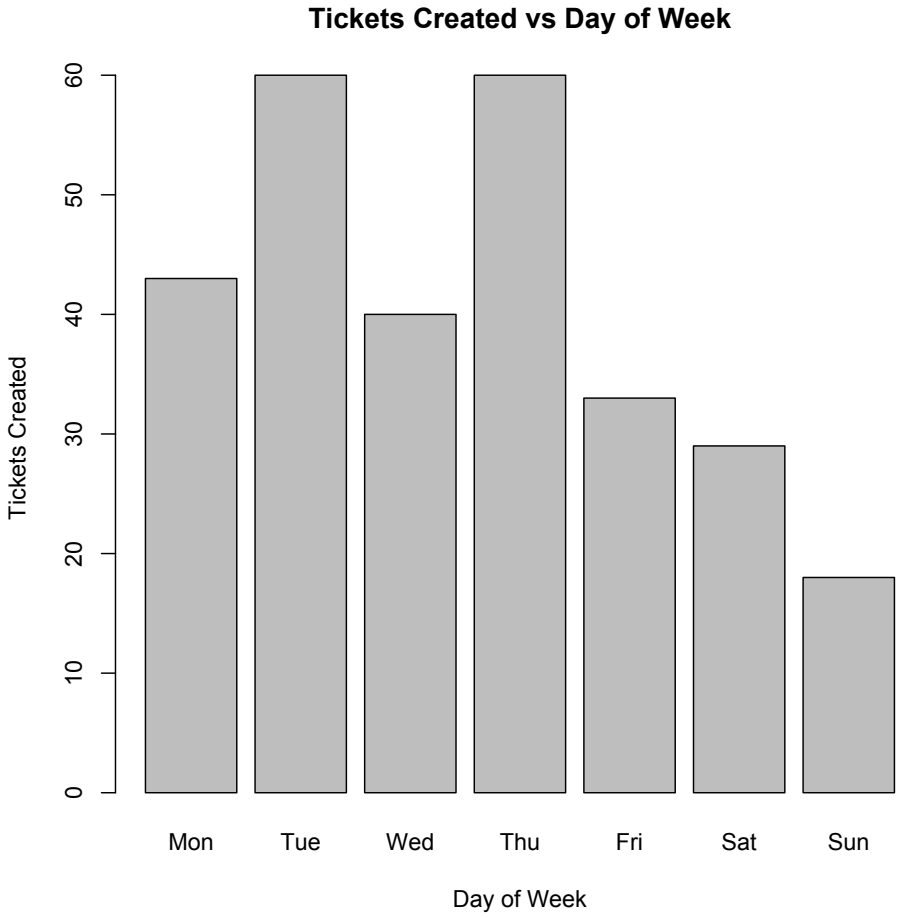

Figure 6: Bar chart of tickets created versus day of the week

Most tickets are opened on Tuesday or Thursday. The least amount of tickets is opened during the weekend.

2.7 Day of the week tickets close

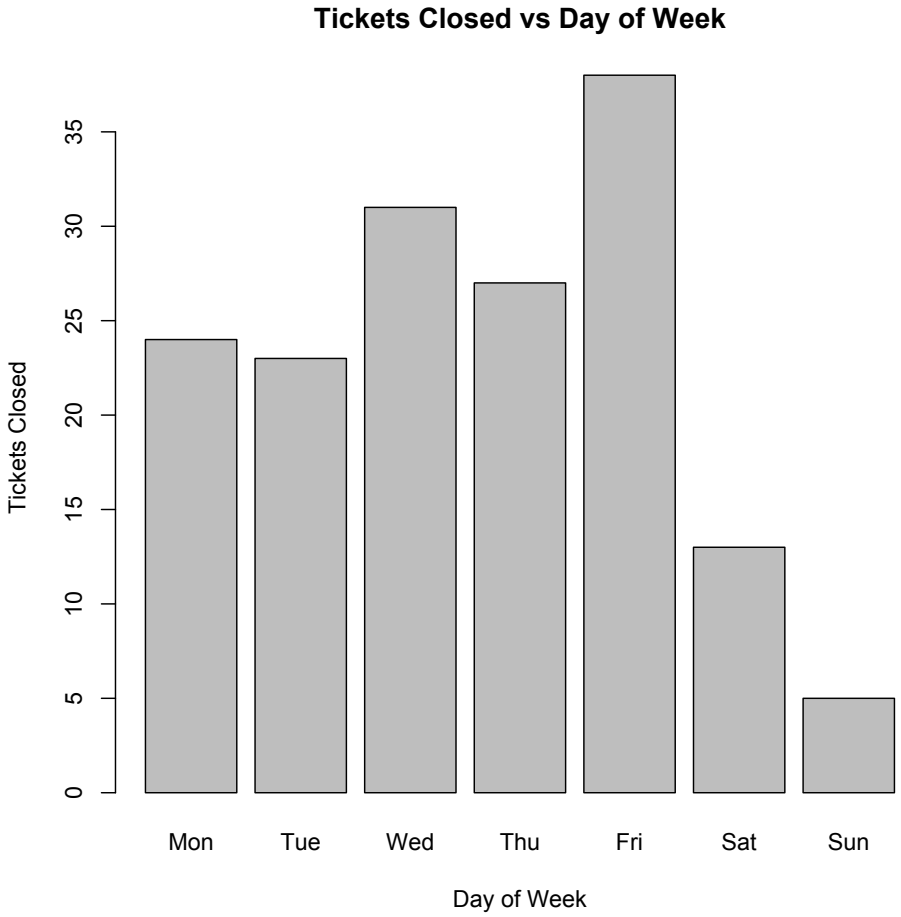

Figure 7: Bar chart of tickets closed versus day of the week

The above bar chart shows interestingly most tickets are resolved on Fridays.

3 Analysis of Ticket Owners

3.1 Ticket lifespan according to owner

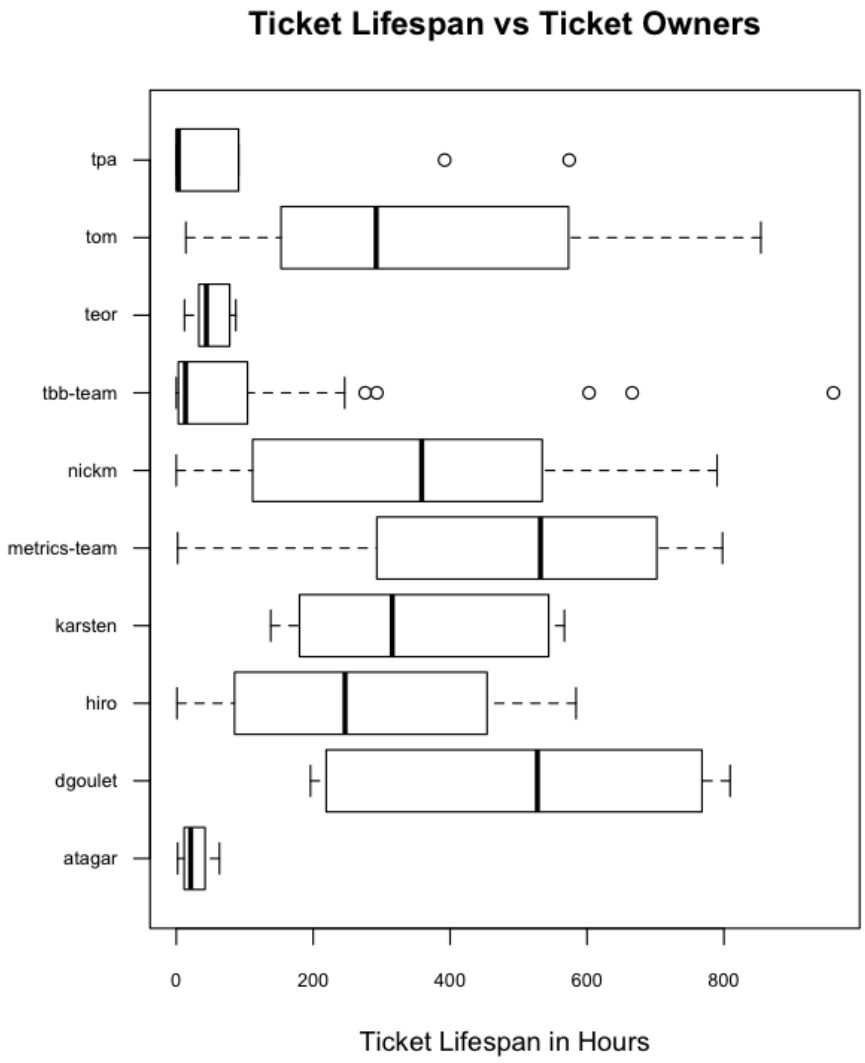

Figure 8: Box plot of owner versus ticket lifespan

This plot is by far the most interesting. The ticket lifespan is plotted for the 10 most active owners, each with a ticket resolution count of 6 or more in January. It is quite remarkable that there are such large variations in ticket lifespan between different owners. It is not clear why some owners have significantly longer ticket resolution time than others. Perhaps, it is

because they were handling more challenging problems or that they were assigned tickets there were not in their area of expertise. To answer this question, more data is needed to categorize and analyze the type of tickets that are assigned to each owner. The answer may provide important clues to optimize the ticket triaging system by assigning tickets to owners who solve certain type of tickets quickly. This may result in a more efficient ticketing system, resolving more tickets in the same period. The percentage of open (new) tickets shown in the above pie chart (Fig.1) should decrease if ticket resolution efficiency increases.

3.2 Composition of owner's ticket queue

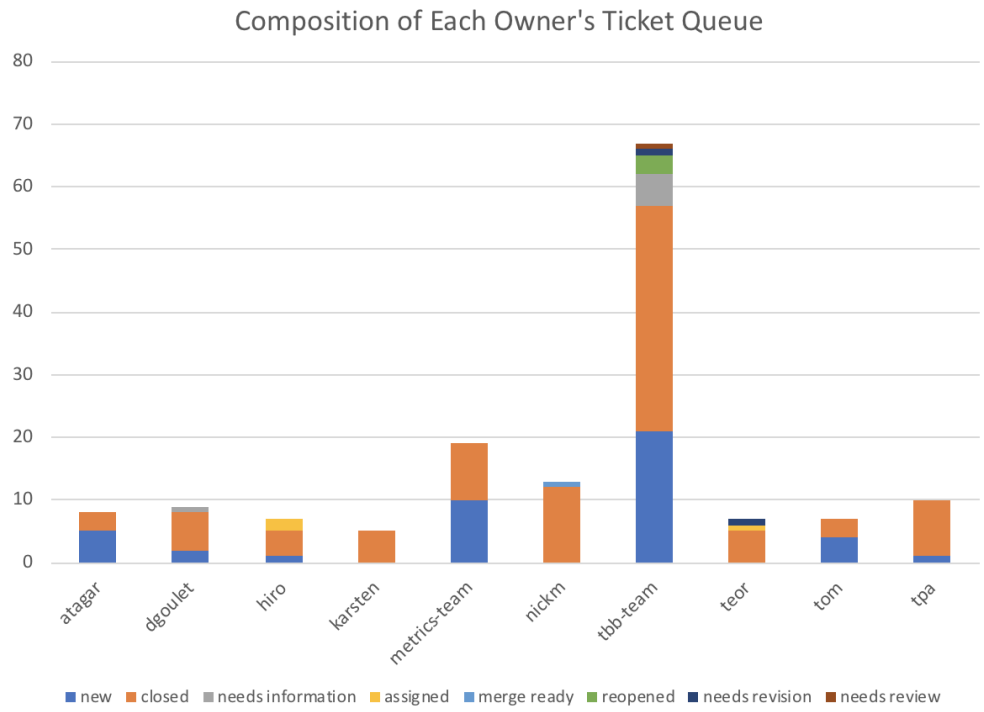

Figure 9: Stacked bar chart displays composition of each owner's ticket queue

The above stacked bar chart illustrates the status of tickets for 10 most active ticket owners. The blue segment indicates the proportion of new (open) tickets associated with a particular owner. Some owners have higher backlog (new tickets) than others, indicating more resources may be assigned to a busy team to reduce the backlog.

4 Conclusion

A set of metrics and plots were developed to analyze various aspect of Tor bug tracking system. These summary statistics and plots can be incorporated into a dashboard to monitor the status of Tor and its bug tracking system in realtime. A change in these statistics could be indicative of a potential problem in the underlying Tor software or the bug tracking process itself. This provides additional information that will help to diagnose and gain insight into problems that may arise.

A secondary goal of this analysis is to objectively uncover the strength of each ticket owner. This information can be used to improve efficiency of the bug resolution by classifying and assigning a ticket to the owner who is most adept at handling a particular type of tickets.

In one month of Trac data, there were 34 owners and 320 tickets. Ticket lifespans (the time it took to resolve a ticket) were manually calculated, then analyzed to evaluate the efficiency of ticket resolution. A preliminary analysis was done and wide variation in the resolution time for different owners were noted (see Fig. 8). However, there is insufficient data to further evaluate which ticket type an owner is most adept at handling. Answering this important question requires extracting additional data, possibly an entire year of Trac ticket history from the database. This requires access to Trac's database using its Python API, as it is too labor intensive to attempt this extraction manually. R will be used for analysis.

Additional monitoring metrics and improved ticket triaging will help to reduce time required to identify and fix issues. Faster diagnostics and shorter ticket resolution would ultimately contribute to a better Tor user experience.