**Protocol analysis approach**

**Problem**

1. Univariate analysis [TODO: give reasons why we're doing these analyses]
   * + How many unique IP-adresses (locations) made a request?
     + From which countries do they probably\* originate?
     + How many IP-adressess accessed the site more than once?
     + Percentage of direct traffic? (“-” in origin variable) / direct vs. internal vs. referral percentages (optional)
     + What is the average (mean,median) times an IP address accessed the site?
     + What type of device was mostly used (mobile+tablet/desktop)?
     + How many visits actually succeeded (vs. failure) to enter the page?
     + Which pages resulted in errors (e.g. page not found)?
     + What were the most requested pages?
     + How many requests did the blog get? (Or the percentage...)
     + What was the number of request for pages from humans?
     + What was the average (mean/median) object\_size in kb?
     + How many documents were downloaded (e.g. PDF)?

From the total of 5000 observations, 2639 (53%) were from unique Ip addresses. 422/5000 (8%) were for a page (a htm, html or php file), excluding a robot or crawler. 311 were unique IP Adresses. 111 Ip-adressess requested more than once.

|  |  |
| --- | --- |
|  | Total (n=5000) |
| Total request from IP addresses, n (%)=422 | 422 (8) |
| Requests of unique Ip addresses (n= 311) | 311 (6) |
| Requests of Ip addresses >1 (111) | 111 |
| Country of origin, n (%)  United States  Belgium  Germany  China  Netherlands  Other countries | 114 (36)  91 (29)  17 (5)  15 (5)  13 (4)  61 (19) |
|  |  |
|  |  |
|  |  |

1. Questions for increase over time
   * + Did the visits increase over time? More visits in April compared to March etc.?
     + In which months most of the visits are performed? (Should we look within this month?)
     + Are there more visits in the weekend compared to the week?
     + Which day (Monday, Tuesday etc.) is most attractive for visitors?
     + Are there more visits in the morning, afternoon compared to the evening?  
       (From the perspective of the server; maybe also perspective of user if we find the time.)
2. Choose a variable and see whether it changes over time [TODO: use these to make tentative recommendations]
   * + Let’s choose number of ‘success rate’ over time. In which month/day/hour is succes rate highest? [Off course it is ok if you have other preference.]
     + How does country of origin ratios (or totals) evolve over time?
     + Evolution of mobile/desktop traffic over time.
     + Pages/visit over time (pages/ip/day)

**Methods**

How are variables defined (character, numeric, date)? For the univariate analysis we will use percentages, numbers, means (SD) or medians (IQR) from variables, depending on normal distribution. (Should we also describe how we converted everything? I think not. That will be in the .rda file. ).

For secondary analyses we will create a day variable classified in morning (7:00-12:00 ), afternoon (12:00-18h) and evening/night(18h-7h) variable and plot numbers over

For the final analysis we defined succesrate of visiting a page as “succeeded” (yes,no)

Direct: you enter the site diretly. Internal+ already in the site and click further. Local 🡪 via the ugent.be for example, search 🡪 from search engines: google, yahoo and bing.

**Results**

The MMA website was visited 5000 times by n= different IP addresses. N= , %, which originated mostly from country y(n=,%) and country x(n=,%) and they were all from the same time zone (200). N, % of the IP addresses was used more than once used to visit the website. Most requests were images (n=, %), only (n=, 10%) of the requests were pages. Of the total requests n= (%) actually succeeded to enter the page.

We did (not) see an increase of visits over months (see plot .). Most visits were made in June (n, %). …… etc.

**Conclusion**

**?**

**Discussion**

* use origin + ip to determine user flows? (landing / exit pages)
* we dint take into account that within a month days might not be equal. For instance there can be more Mondays than tuesdays.