An analysis of TWiki performance following computer simulation of overuse.

Peter L Jones
IT Department, CERN
Geneva 23 CH 1211
Switzerland
Peter.L.Jones@cern.ch

Nils Høimyr
IT Department, CERN
Geneva 23 CH 1211
Switzerland
Nils.Hoimyr@cern.ch

Abstract

The web collaboration platform TWiki (TWiki.org) was installed at CERN by IT-PS following a request for a Wiki from group of software developers. Since then, TWiki has grown in popularity and today the statistics (Jones, 2008) shows more than 4000 registered editors and over 40000 topics. In order to guarantee acceptable performance for the future, a risk assessment would help to identify any bottlenecks that the system may come across. In particular, it may prove difficult for the current file system to cope with the increasing number of topics. Simulating future use can help forecast when such difficulties would arise.

1 Introduction

Wiki pages are like traditional web pages but with one major difference, the user has the possibility to edit the page directly from the browser. For organisations like CERN, this allows collaborations where users are in different buildings, cities or even countries a convenient way of working together on projects and documents.

TWiki, which is just one of many available Wikis (Wikipedia), is a structured Wiki based on resources that are readily available at CERN. TWiki runs on Linux machines with the Apache web server and is written in Perl, the extensively used scripting language. The tool quickly became popular after its introduction in 2003 and as figure 1 shows the number of topics has grown at a steady rate over the past two years.

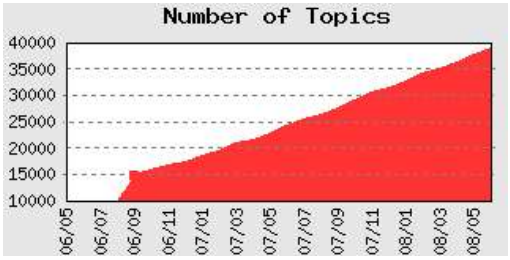


Figure 1. Growth of the total number of TWiki files.

TWiki uses a flat-file back-end and documents for each web (a collaboration area) are all stored in one directory for that web.

Currently there are over 120 webs and these hold between 30 and 9000 topics each. On top of this, TWiki uses the RCS version control system which means that for each topic there is a corresponding history file therefore doubling the number of files in each directory. The data is stored in AFS (OpenAFS) which has a maximum file count limitation of 64k per directory, although this number can be lower depending on length of filename. At the end of June 2008, there were a total of 38554 TWiki topics and the largest TWiki webs, for the CMS and Atlas experiments, had 8755 and 7128 topics respectively.

The number of topics is growing steadily and this paper discusses future performance by predicting how TWiki at CERN will cope assuming the current configuration and conditions.

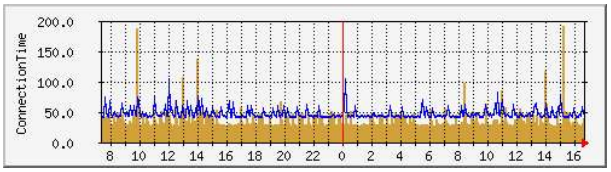
2 Current TWiki performance

The main actions performed by TWiki users are view topic, edit topic and search topic. For the month of June 2008 the statistics show that there were:

- 1367996 total actions
- 30331 monthly updates
- 1291519 topic views
- 7581 searches

Users mostly connect to TWiki to read documents (95% of actions during June 2008 were view actions)

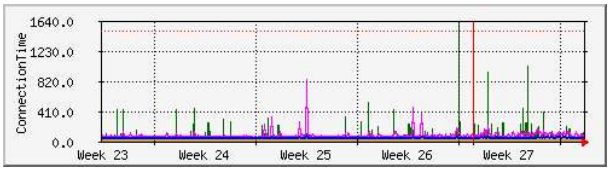
MRTG (Oetike T. 2008) is used to monitor TWiki response times and the graph of Figure 2 shows the average 5 minute page load times for the 18 topics in the Default web (blue line) and the much larger Atlas web (yellow area).



	Max	Average	Current
DefaultWeb:	191.0	36.0	38.0
Atlas:	104.0	47.0	43.0

Figure 2. Daily 5 minute page load averages.

There is little change in the performance figure over a longer period of time as the monthly graph of figure 3 shows.



	Max	Average	Current
DefaultWeb:	1607.0	35.0	42.0
Atlas:	823.0	43.0	47.0

Figure 3. Monthly 2 hour page load averages.

Apart from the odd exception, performance is stable over a number of weeks and there is little difference in response times between the small and large webs.

There is no industry standard for the response time of a web page; however Akamai and JupiterResearch (Akamai 2006) identified 4 seconds as a threshold of user acceptability. The current load times are less than half a second (the figures shown are in 100ths of a second) and so are well below this threshold.

Figure 4 shows the growth in number of topics of the largest web (CMS) over the last nine months. On average there were 326 new topics each month.

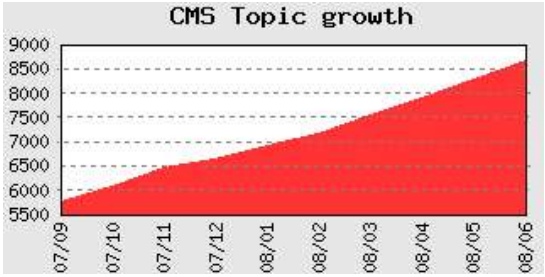


Figure 4. The number of topics in the CMS web.

3 Method of simulation

The simulation is made under similar conditions to the production version of TWiki. This means creating another instance of TWiki in AFS and using the same Linux servers. TWiki is browser driven but for simulation purposes, non-interactive scripts are used to simulate user action with authentication disabled for the test period.

A new empty TWiki web is populated with new topics at regular intervals. The time taken to view, save, search and index these new topics is then monitored.

4 Implementation

In general use, TWiki documents are read, edited and searched for by web browsers using the http protocol.

Figure 5 shows an extract of the script that uses the **wget** command that in turn uses the http protocol to create 20 new topics.

```
#Create 20 new topics
#
$threshold = $count +20;
while($count < $threshold){
   $newtopic = "NewTopic".$count;
   `wget -qO /dev/null 'https://…/twiki-beta
/bin/save/TestWebOne/$newtopic?action_quietsave=1\&
originalrev\&text="This is a new topic number $count"'`;
   $count = $count +1;
}
```

Figure 5.Part of script that creates new topics via http.

Figure 6 is an extract from the script that calculates the time it takes to view the WebHome page of the test web by using the Unix **time** command.

```
/usr/bin/time -f %e -o betaView.out
wget --timeout 20 --tries 1 -qO /dev/null
https://twiki.cern.ch/twiki-
beta/bin/view/TestWebOne/WebHome
```

Figure 6.Obtaining the response time for a web page.

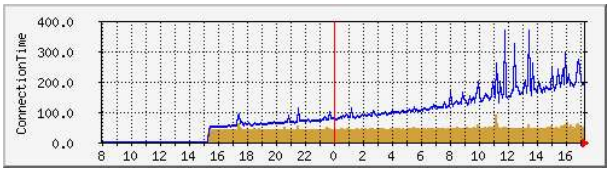
Similar scripts were created for save and search actions and the results fed into MRTG

All scripts are executed every 5 minutes via a crontab entry.

5 Results

5.1 Results for View and Search

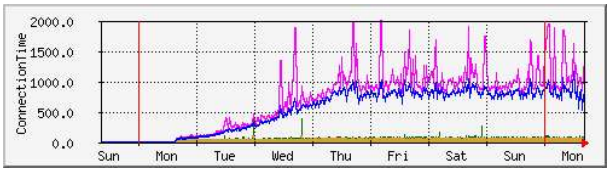
At the end of the first day of simulation 5760 new topics had been created. The blue line representing the search action shows that the time taken for a search grows as the number of topics increases.



	Max	Average	Current
View:	91.0	46.0	48.0
Search:	368.0	115.0	197.0

Figure 7. MRTG daily graph of view and search.

The time taken to load a page for viewing remains stable. Response times for search actions grow linearly until Thursday evening.

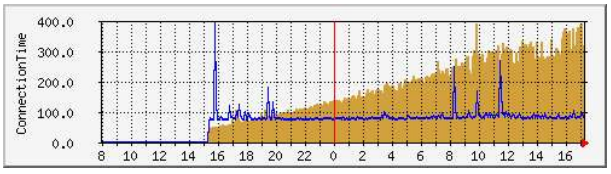


	Max	Average	Current
View:	409.0	60.0	63.0
Search:	2000.0	625.0	749.0

Figure 8. MRTG weekly graph of view and search.

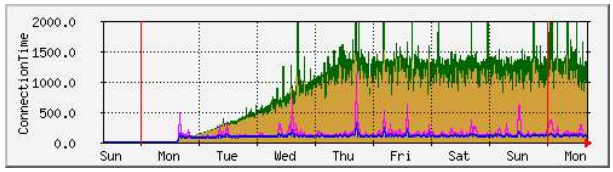
5.2 Results for Save and Index

Figure 8 shows the results of index and save response times after the first day of monitoring. The yellow area representing the index also shows linear growth, while the time taken to save documents remains stable.



	Max	Average	Current
Web Index:	392.0	190.0	302.0
Save:	395.0	84.0	79.0

Figure 9. MRTG daily graph of index and save.



	Max	Average	Current
Web Index:	2000.0	849.0	998.0
Save:	1111.0	100.0	92.0

Figure 10. MRTG weekly graph of index and save.

At the end of the 4th day the script failed to create new topics. The AFS directory could no longer accept new files and figure 11 shows the error sent back from TWiki.

Error saving topic

During save of TestWebOne.NewTopic16105 an error was found by the version control system. Please notify your TWiki administrator.
RCS: failed to create file ./TestWebOne/NewTopic16105.lock: File too large

Go back in your browser and save your changes locally.

Figure 11. No more TWiki topics can be created.

6 Discussion

The experiment shows that the number of topics in a given web has no effect on the response times of the view and save actions. However the search and index actions were apparently affected, the time of response being directly proportional to the number of topics.

The current installation can support around 16000 topics (with RCS that makes 32k files) per directory. Approaching this limit, many topics users will have to wait 9-12 seconds for a response to search and index operations. Following this test, subsequent tests were made using NFS mounted Network-attached storage. The web created on this device supported over 100,000 files (50,000 topics).

The average length of CMS filenames is 25 characters. Repeating the experiment with topics names of this size showed that the web could support up to 15842 topics. Figure 12 shows that given the rate that CMS creates new topics (326/month), the current configuration could cope until the spring of 2010.

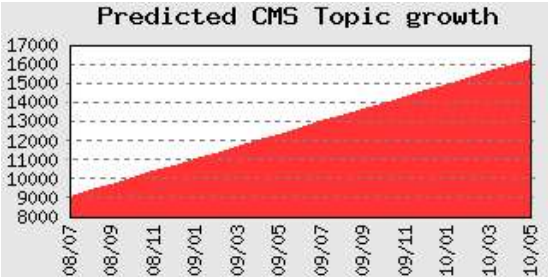


Figure 12. Predicted growth of the CMS web.

7 Conclusions

This paper has discussed how the performance of TWiki would be affected if it continues to be used at the current rate and in particular the impact that very large collaboration webs would have on performance.

A simulation exercise was performed that continually created new topics in a single web and the time taken to perform typical user actions was monitored.

For small webs the current setup is satisfactory with all user operations falling well within acceptable response time thresholds. However users of larger webs will experience slower search and index response times.

The configuration installed on AFS can support all webs for the next 12-18 months but provisions must be made for webs approaching 16000 topics (32000 files). Collaborations do have the option to split their webs or use sub-webs but these solutions have management and usage disadvantages (Thoeny 2008). Other forms of data storage can accept higher numbers of files and so in order to provide acceptable TWiki performance for the future these alternatives should be explored.

References and Bibliography

TWiki.org. (2008), http://www.TWiki.org/

Jones P. (2008), CERN TWiki statistics, https://TWiki.cern.ch/TWiki/bin/view/Main/CERNTWikiStatistics

Wikipedia. . (2008), Wiki Software comparisons, http://en.wikipedia.org/wiki/Comparison_of_wiki_software

CERN. (European Organization for Nuclear Research) (2008), http://user.web.cern.ch/user/cern.html.

OpenAFS. (2008), http://www.openafs.org/

Akamai and Jupiter Research. (2006), http://www.akamai.com/html/about/press/releases/2006/press_110606.html

Thoeny P. (2008), Discussion on TWiki sub-webs, http://TWiki.org/cgi-bin/view/Support/SubWebAdvantagesDisadvantages