Effects of graph embellishments on the perception of system states in mobile monitoring tasks

Frode Eika Sandnes and Kjetil Dyrgrav

Institute of Information Technology
Faculty of Technology, Art and Design
Oslo and Akershus University College of Applied Sciences
Frode-Eika.Sandnes@hioa.no

Abstract. Monitoring of critical systems such as large server parks require suitable visualization tools that draw the attention to essential aspects of system state to ensure continuous operation. Moreover, administration is increasingly performed using mobile devices giving operators more freedom. Mobile device form factors limit the amount of data that can be displayed. This study explores if visual embellishments in data visualizations improve the perception of system states for mobile administrators. The RRDtool time series data visualizer, which is deployed by the Munin monitoring tool, is used as basis and compared to an experimental monitoring tool that employs visual embellishment enhancements. Talk aloud system monitoring sessions were employed with 24 IT-professionals both with and without Munin experience. The results show that the embellishments results in significantly shorter time to interpret the views while they do not significantly help determine the general characteristics of the views. Previous experience with Munin had limited effect on performance. One implications of this work is that embellishments can help focus attention towards important state changes in mobile visualizations.

Keywords: system administration, embellishments, mobile monitoring.

1 Introduction

Control room visualization has received much attention in certain domains such as nuclear reactor monitoring [1], air control [2], airplane cockpits [3], power grid monitoring [4]. Control room visualization methodology has also reached the mainstream though dashboard design, where control room metaphors are used to communicate business states on websites [5].

Monitoring involves detecting abnormal events causing unfavorable changes to the system states. Time is a crucial factor as system state changes that are detected early may cause less damage if a problem is handled early. Unfortunately, many systems are investigated retrospectively using log files after an anomalous event rather than preventively through active monitoring.

Monitoring is increasingly relying on mobile devices giving the system administrators more freedom while increasing their availability and consequently the quality of service as it is perceived by the users. However, system state visualizations on mobile devices is challenging as there is less real estate available for displaying data compared to large control-room display walls [6]. To compensate for smaller displays mobile visualizations need to rely on alternative means to communicate system states. In this study, visual embellishments are explored as a means for enhancing visualizations and drawing the operators' attention to important data. Graphical representation of timeline data is arguably the most common form of visualization estimated to constitute 75 % of all visualizations [7].

This study focuses on three timeline system state parameters that reflect key recourses in computer system, that is, memory use, disk use and processor use. All of these need to be kept below certain levels in order to maintain a healthy system. These parameters are also interrelated. For example, memory data are cached onto the disk. If the memory runs low more data will be cashed, or paged, to disk, consuming more disk space, and even more importantly, the caching consumes or paging processing power and thus affect the processor load.

There is disagreement among experts regarding the approach to information design. Prominent information designers such as McCandless [8] and Holmes [9] design elaborate visual presentations for newspapers and magazines which contrast the minimalistic, minimum ink principles of Tufte [10] and Few [11].

Icons are believed to have an advantage over text in terms of recognition and recall as the human ability to recall images is superior to that for text. Icons can help make decisions more efficiently. However, the meaning of icons must be learned. The visual context of a pictogram significantly strengthens its ability to be correctly perceived [19]. This is the rationale for putting fire extinguisher pictograms on red doors as it assist visual search [12].

Cartoons are examples of a visual media that use exaggerated imagery to effectively communicate meaning through the use of effects such as speech bubbles, lines to indicate motion, explosions, crashes, etc. Another example is to use Zzz as a metaphor for sleeping as Zzz represents the snoring sound [13].

Fire is a powerful image as humans are instinctively adept at responding to fire. Fire is frequently used metaphorically, for example referring to somebody as "on fire" meaning somebody or something is hot without the presences of actual flames. Fire and flames are associated with danger.

This study set out to explore the effect of embellishments for improving the perception of system states as embellishments have been demonstrated to improve long term memory in certain context [14, 15].

1.2 Munin and RRDtool

Current state-of-the-art monitoring tools such as Munin RRDtool have several problems. First, the time series views give misleading information. For example the memory view has an axis which ends at 600 Mb even if the memory device only has a capacity of 512 Mb. The observer may thus be misled to believe that there is more capacity than there really is. Second, the display real estate is ineffectively used. In the previous example, the interval from 512 to 600 constitutes nearly 15% of the view remaining unused. Third, the information provided is inaccurate. The x-axis of many views shows grid lines with 30 hours in total, while the axis date label suggests a 24 hour window. Moreover, the observer has to manually count tick marks to find the time of specific events – a process prone to error and inaccuracies. Four, many views are cluttered. For example, the Munin view shown in Fig. 1a displays 13 parameters, and it is thus hard to identify the individual lines or the overall trend.

2 Method

2.1 Participants

Evaluating and measuring information visualization are challenging [16] and a methodology involving users are needed. Totally 24 participants volunteered for the study. The participants comprised a) IT professionals with practical system monitoring experience and b) network and system administration master students with system monitoring experience from their studies.

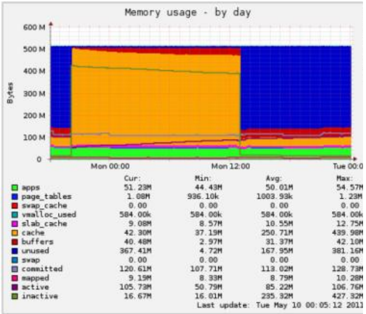
2.2 Equipment

A HTC Desire Andriod smartphone was used for the experiment. It contains a 3.7 inch touch-sensitive display that allows zoom and auto-rotation. The resolution was set to 480 x 800 pixels.

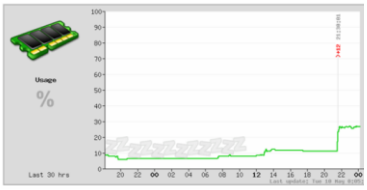
Both the Munin tool with the RRDtool visualization extension and a custom made visualization prototype was used for the experiment. The prototype was written in PHP allowing system commands to be issued on the server to obtain system information about disk, memory and processor usage. The graphics was created with GD (Graphics Draw) PHP library using the Portable Network Graphics format.

2.3 Materials

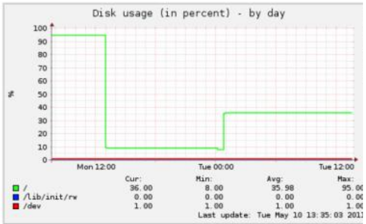
The experiment addressed three monitoring categories, namely memory usage, disk usage and processor utilization. For each category the experiment set out to observe the difference between traditional visualizations without embellishments as utilized by the Munin / RRDtool tool, and enhanced visualizations with embellishments as employed by the prototype tool. The experiment thus comprised six stimuli. To achieve a repeatable and consistent experiment the screenshots where fixed such that all participants were exposed to identical stimuli (see Fig. 1).



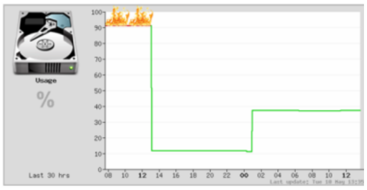
a) Memory use (traditional)



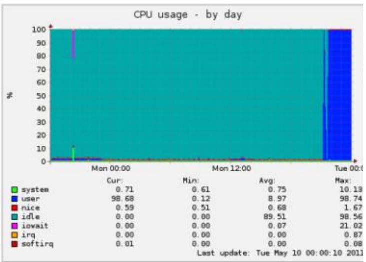
b) Memory use (enhanced)



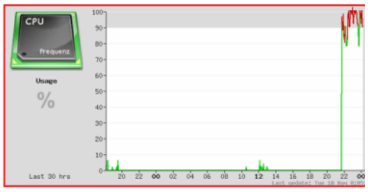
c) Disk use (traditional



d) Disk use (enhanced)



e) Processor use (traditional)



f) Processor use (enhanced)

Fig. 1. The visual stimuli used in the experiments.

The memory scenario starts at low activity during the previous day and lasts for about 16 hours until it increases beyond 10% of full capacity. At 21:30 memory intensive Perl scripts are initiated causing memory use to reach about 30%.

The traditional memory view displays a handful of different memory use categories. During the period of actual low memory use the tool peaks at a very high level due to high cache use (orange). However, cache use does not reflect actual memory use and is somewhat misleading.

The enhanced view is simpler and displays overall memory use in percentage of total capacity. Sleeping cartoon symbols (Zzz) are used to indicate low activity and a red arrow to emphasize the moderate step up from 10% to 30% at 21:30.

The disk scenario starts critically at 8:00 the previous day with approximately 95% of the disk space used. At 13:00 large files are deleted such that only approximately 11% of the disk space is used. Just before midnight new files are created with dd such that disk use increases to 36%.

The traditional view shows disk usage as a simple line graph. The enhanced view employs a red line and cartoon flame embellishments to attract the observers attention to the critical disk levels. No other imagery or embellishments are used in this graph as the step up from 11% to 36% is considered sufficiently noticeable and uncritical.

The processor scenario comprises a timeline where the processor is idle most of the time apart from a few occasional small spikes indicating some system activity. Towards the end of the timeline, that is, at 21:30, the processor usage enters a critical state as it spikes close to 100% due to a processor intensive Perl script.

The traditional view uses different colors for the different classes of processes. It is therefore easy to observe that the first small spike is reflecting system activity (green) and the critical spike towards the end is user activity (blue). Turquoise is used to indicate idle processors. Again the enhanced view shows overall processor use with a red line to indicate the critical state as well as a red line around the entire view.

2.4 Procedure

The experiments were conducted individually. Each participant was sitting down in front of a desk. Steps were taken to prevent glare or reflections from the windows and the light fittings. The experiment was balanced to compensate for any learning effects by presenting the six stimuli in different orders to the participants. A digital voice recorder was placed in front of the participants to capture the dialogue. The participants were informed that the recordings were deleted after analysis.

First, the participants provided demographic information. They were also asked if they had prior experience with the Munin system monitoring tool. Next, they were informed of the structure of the experiment. A talk-aloud methodology was employed and participants were thus asked to describe what they observed, their thoughts and impressions during the experiment.

2.5 Observations and measurements

Time measurements and error rates were measured. For each scenario two success criteria were set up and the time to reach these criteria were measured by analyzing the time interval from the beginning of each scenario was shown to the point where the participant uttered the phrases signaling approximately correct graph interpretations. The following success criteria were devised:

Purpose: the participant identifies the domain of the graph, that is, the type of device that is monitored and the type of data that is measured and displayed (percentage or Mb of memory use, percentage disk use).

Timeline: the participants understand the time-span (24 hours), time-scale and time indication on the graph (trend changes at 11:00, 13:00. 20:00 and 21:30).

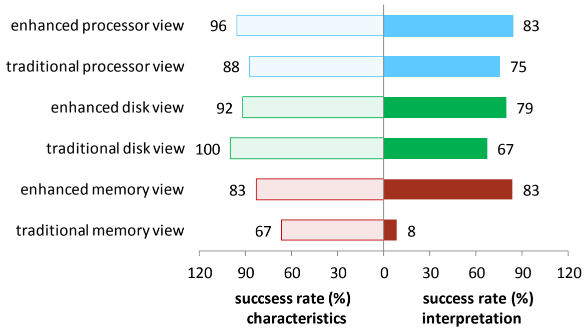


Fig. 2. Success rate in identifying the general characteristics (left) and interpreting (right) the view (100% equals 24).

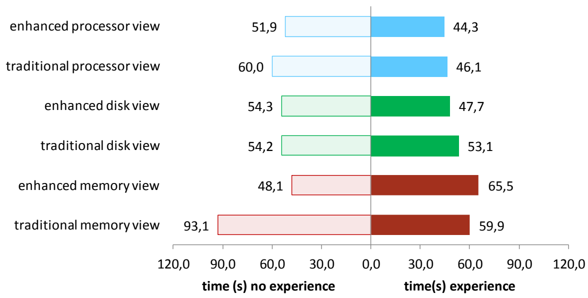


Fig. 3. Mean time identifying general characteristics of the traditional and enhanced memory view (red), disk view (green) and processor view (blue), for participants with and without Munin experience.

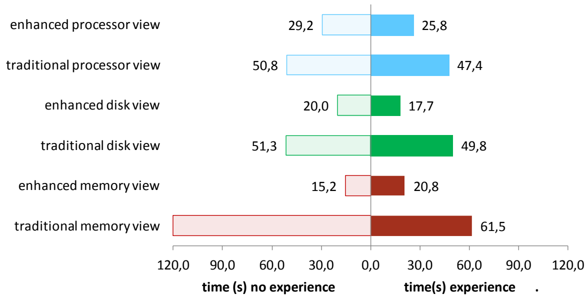


Fig. 4. Mean time interpreting the traditional and enhanced memory view (red), disk view (green) and processor view (blue), for participants with and without Munin experience.

In particular, to what degree are participants able to comprehend the level of detail down to the 30 minute resolution indicated on the graphs (e.g. 21:30).

Interpretation: the participants are able to interpret the trends in the graph and the implications for system state, that is, low, medium, high and critical values. That is, disk usage is critical at the beginning, then low, followed by a small step up, and memory use is first very low with a few spikes then with a critical high spike. Success rates where thus computed as the rate of criteria identified by the participants.

3 Results

The results shows that the associations of the embellishments were reflected in the participants' comments, that is, the most frequent expressions referred to the metaphors sleep (low, idle, calm, quiet), fire (burning, flames) and into-the-roof (high, critical, max, up).

Fig. 2 shows the success rates in identifying the general characteristics of the views (right) and the success rate in interpreting the views (left). Identification of general characteristics of the memory views was associated with the lowest success rates while the disk view was associated with the highest success rate. The general characteristics of the traditional view had the lowest success rate for both the memory test (16 out of 24) and the processor test. All participants identified the general characteristics of the traditional disk view, while two participants did not identify the general characteristics of the enhanced view.

The success rate in interpreting the views (left) is consistently higher for the enhanced views for all scenarios. The traditional memory view has a noticeably low score as only two participants successful interpreted the view.

Fig. 3 shows the mean response time for identifying general characteristics of the three views for participants without (left) and with Munin experience (right). The mean times for the traditional views are consistently longer for all views. The difference is largest for the memory scenario and smallest for the disk scenario. Moreover, the observations are consistent with the success rates shown in Fig. 2 as the memory views are associated with the longest response times and the disk views associated with the shortest response times.

As expected, participants with Munin experience identified the general characteristics of the traditional views faster than the enhanced views, although the differences are marginal. However, the difference between the traditional and enhanced tool is very large for participants with no experience with Munin, as the response time with the traditional view is nearly twice as long as with the enhanced view. Although there is a difference between the traditional and enhanced memory views the difference is not significant ($t = 1.51$; $p > .154$; $df = 13$).

The results for the identifying characteristics of the disk view suggest that there is very little effect of prior experience with Munin, and there is very little effect of enhancements. Consequently, the difference between the traditional and enhanced disk view is not significantly ($t = 1.16$; $p > .258$; $df = 21$).

The results for determining the characteristics of the processor view shows that the experienced participants performed better with both the traditional and enhanced views, and both the experienced and inexperienced participants performed better with the enhanced view compared to the traditional view. However, the differences were not significant ($t = 1.78$; $p > .08$; $df = 20$).

Fig. 4 shows the mean response time for interpreting three traditional and enhanced views for participants with and without Munin experience. Note that only two experienced participants completed the task successfully and the mean time thus exceeds the boundaries of the chart. However, the experienced users took nearly three times longer to interpret the traditional memory view than the enhanced view. The response time for interpreting the traditional disk view is more than twice that for the enhanced view and the difference is significant ($t = 5.09$; $p < .0.01$; $df = 15$).

The results for interpreting the processor view show that both groups perform much faster with the enhanced view than the traditional view and this difference is statistically significant ($t = 3.3$; $p < 0.04$; $df = 15$). There experienced participants also perform marginally better than the inexperienced participants.

4 Discussion

The experiment revealed that 18 out of 24 participants mentioned the term critical when seeing the flames as opposed to 5 participants identifying the through-the-roof method. Moreover, participants needed a mean time of 19.1 seconds to articulate words related to the flame as opposed to 26.5 seconds for through-the-roof. The data thus confirm that fire imagery is more closely associated with critical states.

The red frame was mentioned by only 5 of 24 participants. Despite occupying a substantial amount of real estate, it may not have received attention due to the other visual elements. Similarly the line color was mentioned by only 8 of 24 participants. One explanation is that it was partially hidden behind the flames, as it was only mentioned by 2 of 24 participants for the disk scenario.

Through-the-roof was a frequent phrase uttered by 14 of 24 participants. The snoring embellishment was not associated with the term snoring. Instead, 19 of 24 participants referred to sleeping suggesting that the participants were familiar with the cartoon sleep metaphor. Furthermore, 20 of 24 participants successfully completed this scenario at the shortest mean time (17.4 seconds) for all scenarios. The snoring embellishment is thus likely to positively impact the perception of low activity states.

The traditional view seemed ineffective in communicating memory state as only two participants managed to recognize the memory usage. This may suggest that either the success criterion was biased against the traditional view, or that interpretation of memory usage in Linux is difficult.

Note that the time recorded participants needed to complete the tasks were longer than due to the talk aloud process. This is also a significant source of error as participants may have different speaking styles, verbosity and speech rates. In fact, some participants commented for well over two minutes, including irrelevant aspects of the displays. Moreover, some participants interpreted the task incorrectly or with a wider scope than intended. The talk-aloud methodology was challenging due to mumbling, hesitations and pauses in the middle of sentences.

The differences in experience did not appear as an important factor. However, this could be linked to the fact that the categorization according to experience was based on the participants self reporting and hence not completely objective.

This study suggests that monitoring visualization could benefit from embellishments. Despite warnings against data-interfering embellishments, critical situations where risk is a considerable factor, the discovery of data state should outweigh the desire to render the data accurately.

5 Conclusions

Alternative approaches for enhancing the visual expression of visualized data in system monitoring have been identified though an alternative monitoring tool running in a mobile web browser. It was compared to the commonly used Munin system monitoring visualization tool through three scenarios, namely memory, disk and processor use. Talk aloud sessions were conducted. No significant response time differences were observed in basic graph usage the time difference between the traditional display and the enhanced display, while significant differences were found for monitoring specific usage. The experiment established that use of embellishments significantly improved the perception and interpretation of the state of visualized data.

References

1 Kim, S. K., Suh, S. M., Jang, G. S. et al. Empirical research on an ecological interface design for improving situation awareness of operators in an advanced control room. Nuclear Engineering and Design 253: 226-237, 2012.
2 Rozzi, S., Amaldi, P., Wong, W. and Field, B. Operational potential for 3D displays in air traffic control. In proceedings of ECCE '07: Proceedings of the 14th European conference on Cognitive ergonomics, ACM press, pp. 179-183, 2007.
3 Pinder, S.D. and Davies, T.C. Interface Design for an Aircraft Thrust and Braking. In proceedings of OZCHI 2006, ACM Press, pp. 103-110, 2006.
4 Haigh, D. and Adams, W. Why 64 bit computers in the control room, Computing & Control Engineering Journal 7(6): 277-280, 1996.
5 Few, S. Information Dashboard Design, O'Reilly Media, 2006.
6 Chittaro, L. Visualizing information on mobile devices. Computer 39:40–45, 2006.
7 Tufte, E. R. The Visual Display of Quanitative Information. Graphics Press LLC, 2009.
8 McCandless, D. Visual Miscellaneum, Harper Design, 2012.
9 Holmes, N. Designer's Guide to Creating Charts and Diagrams, Watson-Guptill, 1991.
10 Tufte, R. R. Envisioning Information. Graphics Press LLC, 1990.
11 Few, S. Show Me the Numbers - Designing Tables and Graphs to Enlighten. Analytics Press, 2004.
12 Abdullah, R. and Hubner. R. Pictograms, Icons and Signs. Thames and Hudson, second edition, 2006.
13 Healey, C.G. Choosing effective colours for data visualization. In proceedings of the Seventh Annual IEEE Visualization '96, IEEE CS Press, pp. 263-270, 1996.
14 Bateman, S., Mandryk, R. L, Gutwin, C., Genest, A., McDine, S. and Brooks, C. Useful junk?: the effects of visual embellishment on comprehension and memorability of charts. In proceedings of the 28th international conference on Human factors in computing systems, CHI '10, ACM Press, pp. 2573–2582, 2010.
15 Borgo, R., Abdul-Rahman, A., Mohamed, F., Grant, P. W., Reppa, I., Floridi, L. and Chen, M. An Empirical Study on Using Visual Embellishments in Visualization, IEEE Transactions on Visualization and Computer Graphics 18(12): 2759-2768, 2012.
16 Plaisant, C. The challenge of information visualization evaluation. In Proceedings of the working conference on Advanced visual interfaces, AVI '04, ACM Press, pp. 109–116, 2004.