Comments from Examiner 1 :-

Detailed Comments on the Ph.D. Thesis
“Operating System Auditing and Monitoring”
by Wu Yongzheng (A0002223H)

General Comments on the Thesis

Overall I believe that this thesis presents novel and significant research at a standard comparable to Ph.D. theses from the top international computer science departments. The thesis presents a nice mix of results relevant to both systems developers and system users. The large set of quality publications produced by the candidate is a further indication of the quality of the work.

Most of the research was targeted to the Microsoft Windows platform, and while the thesis repeatedly claims that the results can be generalized to other platforms (which I accept), I wonder to what extent the research addresses problems that are (as Fred Brooks would say) “accidental” to Windows rather than “essential” to systems in general. For instance, the reliance on DLL sharing in Windows is an artifact of the need to conserve storage, but this need arguably no longer exists, and platforms that don’t employ sharing like this (such as Mac OS X) avoid many of the problems that sharing creates and are solved by the research in this thesis.

The presentation of the thesis is well organized, and the writing is good. The thesis presumably was constructed out of the publications that were produced, and the candidate has done a nice job of hiding the “seams” and ensuring that the end product is cohesive. There are numerous minor lapses in English usage (too numerous and minor to pedantically enumerate), particularly missing or duplicated or spurious articles and prepositions, so a final proofreading would be helpful. (A few examples are included in the detailed list below.) In addition, I personally find the use of citations as nouns to be jarring; a sentence should still make sense and be grammatically correct if the citations were removed. And in some places the placement of figures and tables seems haphazard, with some figures being referenced many pages before (such as Figure 5.4) or after (such as Figure 5.18) their placement, and others being referenced in an order different from the order of placement; assuming the text was typeset in Latex, in most cases the figures and tables should be placed immediately following their first place of reference.

List of Amendments to Be Made in the Thesis

Chapter 1: Provide references for principles, systems and techniques that are mentioned (such as RootKit, DTrace, SystemTap, the principle of least privilege, etc.).

Section 1.2: Briefly clarify what key *novel ideas* are embodied in the infrastructures and visualizations.

p.13 middle: “mode relevant” -> “more relevant”

p.19 top: “printk” -> “printf”

p.27 middle: Spell out IDS the first time it is used.

p.28 item 2: Briefly explain the difference between the four kinds of IDs.

p.33 top: There is spurious text at the end of the first paragraph.

p.34 Figure 3.1: Is it necessary to cast the “event” variable in the switch condition?

p.35 top: “as lone as” -> “as long as”

p.38 Section 3.1.3.2: The example implies that some predefined pid variables (and possibly other variables) are available, but such variables aren’t mentioned or described anywhere.

p.39 bottom: “worse case” -> “worst case”

p.40 item 5: “containing the which” -> “containing the file which”. The item also has an extra period at the end.

p.40 Table 3.1: Mention the units in the caption too.

p.41 top: There seem to be missing or extra words in “Systrace on the other as it does access”.

Section 3.1: Somewhere in this section, for completeness provide a brief description of how LBox works (such as a block diagram) and a brief description of the kernel modifications needed (how many or which source files, etc.). It might also be useful to provide a complete syntax or API for LBox as an appendix to the thesis. Note that Section 3.2.2.1 presents much more detail for WinResMon than is provided in corresponding fashion for LBox.

p.52: Mentions Appendix 2, but there is no Appendix 2 (or Appendix 1) in the thesis.

p.55 footnote: “after x step 4” -> “after step 4”

p.70 Section 4.3.1.1: Provide hardware and OS details for the experimental setup.

p.77 Section 4.3.2.1: The section presents some conclusions but no detailed results like previous sections have provided.

p.78 Figure 4.4: Specify the units for the x-axis.

p.79 Figure 4.5(b): Briefly discuss Figure 4.5(b) in the text, since its meaning is unclear.

p.85: "utilize low resources" -> "utilize resources at low rates"

p.94ff.: Somehow it needs to be clarified explicitly early in Section 5.1.2 that the graphs depict dynamic runtime dependencies and not static structural dependencies.

p.99: There are a few typos in the bottom paragraph. In addition, the "space limitations" mentioned there (presumably from a conference version of the text) don't really apply for a thesis.

p.101 Figure 5.4ff.: Provide an explanation in the text for the edge weights appearing in these graphs.

p.120: Clarify how the aggregations are specified or determined.

p.121: Clarify the meaning of the axis barcodes, which are *projections* and/or *aggregations* of some kind.

p.130: Present rather than omit the time-ordered comparison mentioned for cp and xcopy

p.132: Present rather than omit the zoom-in mentioned for the comparison of cp and xcopy.

Chapter 5: Clarify somewhere in this chapter exactly what forms of interaction are supported by the tool.

Sections 6.1-6.2: The amount of spurious and missing words was particularly noticeable in these sections.

p.145: MAC is expanded explicitly to Medium Access Control toward the top of this page and then used implicitly to mean Message Authentication Code starting from the bottom of the page. Resolve this somehow, such as re-expanding MAC to indicate the second sense.

p.158 middle: "of of" -> "of", "we evaluation" -> "we evaluate"