Dear Editor and Reviewers,

We thank the reviewers for their comments. We have carefully revised the paper following these comments.

The major changes include:

1. We follow the 1st reviewer’s suggestion by restricting the scope of this paper to the multitasking mechanism between activities and fragments.

2. We improve the presentation of Section 3 as follows.

(1) To simplify the definition of the semantics of AMASS𝐴𝐶𝑇,𝐼𝐹, we choose to skip the NOH flag in Section 3.1 (as a result, the definition of formal semantics is shortened from over 3 pages to around 1 page). The full semantics of AMASS𝐴𝐶𝑇,𝐼𝐹 is included in the appendix.

(2) Before the technical definition of the formal semantics, we use examples to help the readers get an intuitive understanding of the semantics first.

3. In Section 6, besides the semantics-validation experiments, we also audit the source code of the activity-fragment multitasking mechanism in Android OS to confirm the consistency of the formal semantics to the source code. The manual code audit and automated semantics-validation experiments both increase the confidence in the correctness of the formal semantics.

4. We add discussions on the limitations of the dynamic model extraction in Section 4. The effect of the dynamic model extraction on the performance of ICCBot\_{AMASS} is demonstrated by the experiment results in Tables 9 and 10.

5. We discuss the implications of their work in practice in the conclusion.

~~-------------------------------------~~--------------------------------

Below are the answers to the reviewer’s questions and comments.

**Review 1**

**Q1**. I still have doubts on the claim from the authors about the multi-tasking mechanisms in Android. The authors indicate that all of the high-level multitasking is implemented via (relatively) lower-level inter-component interaction/communication. But, for different types of high-level multitasking mechanisms for different components, the low-level implementation and modelling can be different. For example, the intent flags used for activities can be totally different from those used for starting Services. In addition, Android also supports other types of multi-threading such as AsyncTasks. However, in Section 2, the authors only use multi-tasking to refer to “activities and fragments, and the evolution of the back stack”, which, I think, are only a type of multitasking on Android. I think the authors should better restrict the scope of this paper. Instead of claiming that they are formalizing the multitasking mechanism of Android, they should explicitly specify that they are formalizing multitasking between activities and fragments.

**A**. We restrict the scope of this work to the Android activity-fragment multitasking mechanism. Parts of this paper, including the title, abstract, and introduction, have also been adapted accordingly.

**Review 2**

**Q1**. I thank the authors for adding more details regarding the cross-version difference in the mentioned aspects. However, the current Section 3 is a bit unorganized. The authors should consider organize them into subsections or tables for better readability.

**A**. We organize Section 3.2 (semantics of AMASS\_ACT for the other versions of Android) into subsections.

**Q2**. Please add discussion on the limitation of using dynamic testing for model extraction, and clarify how this could affect the performance of ICCBot\_{AMASS}.

**A**. We have added the discussions on the limitation of dynamic model extraction in Section 4 and clarified that the dynamic model extraction in ICCBot\_{AMASS} is slower than the static model extraction in general. The experiment results for the performance comparison can be found in Section 8 (Table 9-10).

**Review 3**

**Q1**. Although the authors claim to rewrite Section 3 completely, this section is still not understandable. As a consequence, the correctness of the developed formal semantics for Android system is still hard to judge. The authors' response "the precision/correctness of the model extraction is somehow not central to the goal of this paper" somehow sounds very strange to me.

The presentation of the paper, especially those technical parts, is not much improved. (Some examples: Section 3.1 - transition relation, Semantics of AMASS𝐴𝐶𝑇,𝐼𝐹 pp15-19.)

**A**. We improve the presentation of Section 3 as follows.

(1) To simplify the definition of the semantics of AMASS𝐴𝐶𝑇,𝐼𝐹, we choose to skip the NOH flag in Section 3.1 (as a result, the definition of formal semantics is shortened from over 3 pages to around 1 page). The full semantics of AMASS𝐴𝐶𝑇,𝐼𝐹 is included in the appendix.

(2) Before the technical definition of the formal semantics, we use examples to help the readers get an intuitive understanding of the semantics first.

**Q2**. The correctness of the proposed formal model is difficult to judge. The validation of the formal semantics in Section 6 is insufficient.

**A**. In Section 6, besides the semantics-validation experiments, we also audit the source code of the activity-fragment multitasking mechanism in Android OS to confirm the consistency of the formal semantics to the source code.

**Q3**. The authors should discuss what the implications of their work in practice.

**A**. We discuss the implications of the work in practice in the conclusion.

**Q4**. The authors should also discuss whether the proposed formal model is still relevant for later Android versions.

**A**. We define the semantics of AMASS for the latest Android version (i.e. Android 13.0) in Section 3 and also discuss the differences in the semantics for the other versions (i.e. Android 6.0-12.0).