Praxis trip III: Towards Empirical Evaluation

1 Introduction

This latest Praxis visit seeks to gain industrial strength empirical evaluation of NuSPADE. This note was written prior to arriving at Praxis to gain a clearer idea of the evaluation process required[1].

2 Don't Panic!

2.1 NuSPADE weaknesses

NuSPADE tackles program reasoning on both fronts: with automated proof discovery and automated program analysis. Although there is little doubt about the merits of the NuSPADE approach, there are problems in actually making it go in practise. Fundamentally, real SPARK can be quite complex. Building comprehensive tool support, on two fronts, for an entire language is a little ambitious. Shortcuts have been taken.

- **Program analysis -** NuSPADE is capable of discovering program properties for a workable portion of SPARK provided the user generate some additional program description files (essentially extended fdl).
- **Proof planning -** NuSPADE can discover proofs and errors, prompting program analysis if required.
- NuSPADE **as a whole -** The integration of these two components is completely absent. This task has been overshadowed by the more academically relevant challenge of discovering new information.

However, despite such shortcomings, NuSPADE can tackle suitably formatted examples. It is quite possible to translate nearly all of SPARK (excluding the low level aspects) into examples understood by NuSPADE. Although employing NuSPADE can take a little tweaking, it automates the difficult aspects, successfully demonstrating the potential of a fully fledged implementation.

2.2 NuSPADE niche

Another issue is the increasingly small niece (exception freedom) that the NuSPADE system targets. The Examiner now adds type information which has led to the successful discharge of further VCs. Also, the Simplifier has been improved, successfully discharging more VCs.

These success stories have been achieved by techniques in line with the NuSPADE approach. Although not following the NuSPADE architecture, the similarities (strengthen the specification / adopt reasoning heuristics) provide encouraging results. Although this does serve to narrow the exception freedom domain, there are still difficult problems to be found which NuSPADE is likely to make progress with.

3 Tasks

The problem of industrial evaluation of NuSPADE is decomposed into a series of steps:

- **Arranging -** Arranging the input in the appropriate form. These will be isolated packages containing code that faithfully represents (often exactly) a section of a high integrity system. The main difference with these isolated packages will be the local grouping of variables and type and the creation of extended fdl files.

  These isolated packages will focus on the more problematic areas of the high integrity system. Thus these will cover a small percentage of the system as a whole. It is important that this is stressed in any documentation of the results.

  It is important to statistically relate these isolated package to the original code. Comparing VCs proved in the real system with those in the isolated package is likely to be confusing and possibly unfair on the SPARK

*Notes in this series are for ϵ-baked ideas, for $1 \geq \epsilon \geq 0$. Only exceptionally should they be cited or distributed outwith the NuSPADE project (EPSRC GR/R24081).

[1] As it transpired, this process was not entirely successful. However, as described in CS-note , it mutated into something more practical.

tools. Thus here it is reasonable to count the total number of subprograms in the real system and those that appear in the isolated packages. This would only give accurate results if subjecting the entire system to careful scrutiny. This problem can be avoided by working on a per file basis, building up isolated packages gradually and accurately recording their frequency with respect to these files.

Note that it is conceivable that these isolated packages may be innocent and infrequent enough to travel back to Heriot-Watt.

- **Running -** Running the SPARK tools on the isolated package and gathering results. Running the (fragmented) NuSPADE tools on the isolated package and gathering results.

  The proof comparison can be performed on the isolated packages themselves. The SPARK tools will be able to tackle these packages and provide individual results. It will be much more straight forward to compare the proof results of several examples individually rather than many altogether.

- **Tabulate -** Collate the results in a clear tabular form.

4 Production

The key production is the isolated packages and the statistics that relate these to the high integrity system they were extracted from. These isolated packages are then be subjected to analysis as desired. It is important that analysis is performed while the isolated packages are collected to check their quality and correctness and gauge progress.