Four birds with one stone: On the use of time-randomized processors to address the time-predictability, reliability, energy and security challenges of real-time autonomous systems

David Trilla[†,‡], Carles Hernandez[†], Jaume Abella[†], Francisco J. Cazorla[†]
[†]Barcelona Supercomputing Center (BSC)
[‡]Universitat Politècnica de Catalunya (UPC)

I. Introduction

Autonomous Vehicles, in particular cars, are set to be one of the nearest technological breakthroughs of our era. Enabled by the latest advances in machine learning and artificial intelligence techniques, manufacturers find themselves in an arms race to overcome the challenges of designing systems in one of the most complex and holistic domains of IT, which involves aspects as diverse as security, reliability, energy consumption and last but not least, time-predictability. However, this technological revolution might be halted by the important challenges posed by the requirements of critical real-time systems built into them. Up until now, processor's architecture of real-time (critical) systems has been kept simple in the embedded domain in contrast to the advances of the state of the art HPC processor solutions. The reason being the need to provide evidence of functional and non-functional (e.g. timing) correctness. However, the emergence of 'smart' critical software services, e.g. autonomous driving, in critical systems demands higher levels of computing performance that can only be realistically provided by incorporating advanced (and complex) processor features. This creates opposing design principles: keeping a simple and as predictable as possible processor design to facilitate verification of functional and non-functional metrics vs. using high-performance and complex processor designs with some degrees of obfuscation to deliver high computing performance and security.

Time-Randomized Processors (TRP) have recently been proposed to achieve a compromise between both design principles. TRP are an alternative to traditional (deterministic) designs. TRP facilitate timing analysis via the use of statistical/probabilistic techniques, while also show capabilities to effectively tackle the challenges of reliability, security and energy consumption. Due to the statistical properties and probabilistic behavior of these processors, they are able to: (1) Provide statistically founded estimates of real-time tasks' execution times. (2) Expose, detect and bound the magnitude and frequency of power dissipation peaks. (3) Provide probabilistically bounded estimates of the energy cost of task execution. (4) Mask information leaks that compromise security and occur when tasks share hardware resources. (5) Equalize and distribute the load of highly utilized processor features increasing their lifetime. In this contribution, we review the TRP opportunities and show they are a natural fit to fulfill the requirements of autonomous critical systems.

II. Time-Randomized Processors

The design of TRP [1] revolves around controlling the jittery resources of the processor. There exist two main approaches to this: either upper-bounding operations by forcing maximum latency so the time to access a given resource is constant, or, conversely, randomizing the timing behavior to access that resource. There are several processor resources that can be randomized, such as the bus arbitration policy, cache placement, and cache replacement. Figure 1 shows an example design for randomizing cache placement. In

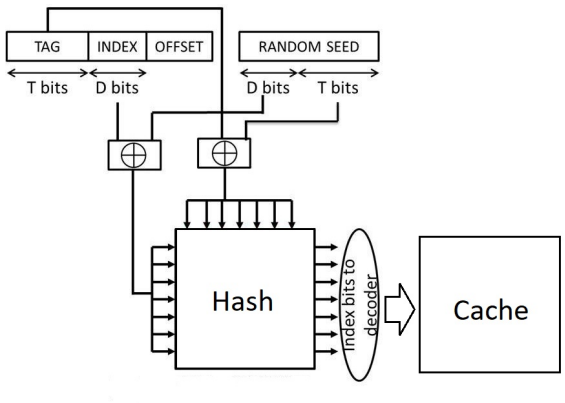

Fig. 1. Design of a randomized cache placement for a TRP. Memory addresses are combined with a random seed to generate a random set index.

tipical deterministic caches, the index bits for determining which set to access are computed through a modulo operation on some address bits. In contrast, a randomized cache uses a hash function combining address bits and a random number to generate a random set number. The implication of this is the randomization of cache access, essentially randomizing interference and decoupling each execution from the previous state of the cache, and randomizing execution times.

III. Time-Predictability

The particular design of TRP allow execution times of tasks to exhibit statistical properties. By upper-bounding and randomizing sources of jitter, each execution of a task is an independent experiment over a sample space that is all the possible execution times. With this property, even if the processor is complex, with a sufficiently large number of measurements we can observe a variety of execution times, and then derive probabilistic estimates on the worst-case execution time (WCET) of tasks. Statistical treatment of the execution time measurements (i.e. Extreme Value Theory (EVT)) allows deriving a mathematically backed estimate of the probability of exceedance of the WCET estimate.

EVT is a mathematical tool that allows, given a set of independent and identically distributed (i.i.d.) measurements, deriving a probabilistic projection of the tail of their distribution thus allowing to predict the worst cases and bound them probabilistically. TRP contribute in giving i.i.d. properties to the execution times, hence enabling the use of EVT to obtain timing predictability on high-performance processors. The complete use of these techniques is usually referred to as Measurement-based probabilistic timing analysis (MBPTA) [2], see Figure 2.

IV. Power Peaks

As more powerful processors are introduced into the real-time domain, the number of unexpected behaviors also increases. One of

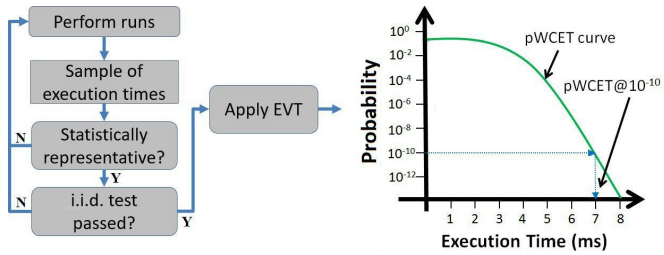

Fig. 2. MBPTA process to generate probabilistic WCET estimates displayed as probabilities of exceeding specific execution times.

the most affected aspects when increasing processor complexity is power dissipation. Power spikes can happen when multiple parts of the processor activate at the same time drawing simultaneously power from the power distribution network. This kind of effects can have a wide variety of effects on the system, from damaging it to decreasing performance through dynamic voltage and frequency scaling, in some cases producing deadline misses and compromising the validation process. For Critical Real-Time Systems this is specially important since they require thorough validation.

The frequency of power dissipation peaks as well as their magnitude is dependent on multiple factors that end users are not likely to be in control. This is specially true as processors get more complex. As more independent circuits are added, the sample space increases in dimensions making it unfeasible for a particular tester to derive the corner cases. Even when those peaks are exposed, if they become a pathological behavior, hardware degradation may diminish the lifetime of the processor.

TRP help in two ways. First, during testing, they expose patterns that power viruses might miss, and second, they break pathological behavior. The key lays in the randomization mechanism. The randomization mechanism allows the tester to explore all the possible power dissipation values as independent experiments, where each execution is a random observation over all the possible combinations of hardware activations. Therefore, TRP make surface behaviors that might typically escape a tester designing a power virus. Similarly, if a pathological case is detected, randomized processors affect each execution independently, therefore the power peak will have lower probability of repeating than in a deterministic processor, therefore increasing lifetime for the hardware.

V. Energy Consumption Estimation

The increasingly powerful processors needed in the embedded domain also bring an increase in energy consumption needs, diminishing battery lifetime and compromising the completion of critical tasks. Critical real-time systems demand thorough validation of their non-functional aspects, including energy consumption, to ensure that critical tasks have enough energy to execute until completion.

Similarly to Time-Predictability, complex processors challenge energy estimation of tasks. Current state of the art models do not provide a solid foundation on which to base worst-case estimates, since they are too abstract and, therefore, have low accuracy, or too specific and not scalable to full task execution. By using TRP and the MBPTA methodology, we enable mathematical soundness for Worst-Case Energy Consumption estimation.

VI. Side-Channel Attacks Resilience

Connected vehicles are one of the goals of the current autonomous revolution. Complex software is being installed in cars and vehicles that will require remote updates, connection to internet, or even third party software. This brings to the embedded domain the security problem that any connected devices face opening up for security intrusions never seen in the automotive or real-time domain.

Side-Channel Attacks (SCA) have recently gained relevance in the security community due to the severe implications in security. SCA are basically security intrusions that use information leaked to the environment to extract sensitive data. There are multiple types of SCA depending on which is the physical domain where the information is being leaked, one of which is Timing Attacks. In particular, Cache-Timing SCA use the information that the access to cache latency leaks to the timing domain. In this kind of attacks, a malicious actor is able to deduct what data is present in the cache by observing the access times. Coupled with input dependent memory accesses, an attacker is able to infer secret data by exploiting the deterministic mapping of caches. One example of this are contention based attacks, where an attacker manipulates the input data to evict victim's data present in cache. If the attacker is able to observe longer execution times, he can deduce which data was the victim using.

TRP break the correlation between the value of input dependent memory accesses and placement in cache by using the randomized hash mechanism. With a unique random seed per process, other processes can not reliably and deterministically evict each others data, therefore breaking the information channel that a malicious actor could use.

VII. Increased Reliability

Future autonomous vehicles and real-time devices are set to integrate the latest technology in processor design while maintaining the request for extended lifetime as the average car, plane and its embedded computers are expected to last longer than the average end-user computing platforms. As processor feature sizes decrease, the impact of variability in transistor increases. This magnifies the impact of aging since small accumulated defects over time can produce failures much faster in smaller feature sizes. This is specially important for critical devices that have a prolonged and continuous usage.

In general, aging effects can be modeled after the usage of the hardware (i.e. the more the hardware circuit is activated, the worse the aging impact, hence the greater the decrease in lifetime). In deterministic systems, biased executions that abuse the usage of certain parts of the hardware can greatly degrade lifetime. For instance, a software that uses statically memory mapped tables, will map those memory accesses to particular sets in the cache. These sets will be used more than its peers, therefore suffering more aging.

Randomized placement in TRP decouples the relation between the virtual indexing value and the final set placement by combining the address with a random seed in a hash function. By doing so, TRP effectively randomize placement across all sets distributing usage across all the cache, therefore increasing set lifetime. Biased executions that are constantly mapped into the same memory region access sets across all the cache independently of the index bits of their virtual address [1].

References

[1] D.Trilla et al., "Aging Assessment and Design Enhancement of Randomized Cache Memories," IEEE TDMR, vol. 17, pp. 32–41, March 2017.

[2] L. Cucu-Grosjean et al., "Measurement-based probabilistic timing analysis for multi-path programs," ECRTS, July 2012.