
To cite this version:

HAL Id: hal-04108237
https://inria.hal.science/hal-04108237
Submitted on 31 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Benjamin Rouxel\textsuperscript{1}, Christopher Brown\textsuperscript{2}, Emad Ebeid\textsuperscript{3}, Kerstin Eder\textsuperscript{4}, Heiko Falk\textsuperscript{5}, Clemens Grelck\textsuperscript{6}, Jesper Holst\textsuperscript{7}, Shashank Jadhav\textsuperscript{8}, Yoann Marquer\textsuperscript{8}, Marcos Martinez De Alejandro\textsuperscript{9}, Kris Nikov\textsuperscript{4}, Ali Sahafi\textsuperscript{3}, Ulrik Pagh Schultz Lundquist\textsuperscript{3}, Adam Seewald\textsuperscript{10}, Vangelis Vassalos\textsuperscript{11}, Simon Wegener\textsuperscript{12}, Olivier Zendra\textsuperscript{*13}

\textsuperscript{1}Unimore, Italy
\textsuperscript{2}University of St.Andrews, United Kingdom
\textsuperscript{3}University of Southern Denmark, Denmark
\textsuperscript{4}University of Bristol, United Kingdom
\textsuperscript{5}Hamburg University of Technology, Germany
\textsuperscript{6}University of Amsterdam, Netherlands
\textsuperscript{7}SkyWatch A/S, Denmark
\textsuperscript{8}University of Luxembourg, Luxembourg
\textsuperscript{9}Thales Alenia Space, Spain
\textsuperscript{10}Yale University, USA
\textsuperscript{11}Irida Labs AE, Greece
\textsuperscript{12}AbsInt Angewandte Informatik GmbH, Germany
\textsuperscript{13}INRIA, University of Rennes, CNRS, IRISA, France

Abstract—Non-functional properties, such as energy, time, and security (ETS) are becoming increasingly important in Cyber-Physical Systems (CPS) programming. This article describes TeamPlay, a research project funded under the EU Horizon 2020 programme between January 2018 and June 2021. TeamPlay aimed to provide the system designer with a toolchain for developing embedded applications where ETS properties are first-class citizens, allowing the developer to reflect directly on energy, time and security properties at the source code level. In this paper we give an overview of the TeamPlay methodology, introduce the challenges and solutions of our approach and summarise the results achieved. Overall, applying our TeamPlay methodology led to an improvement of up to 18% performance and 52% energy usage over traditional approaches.

Index Terms—Energy, Real-Time, Security, System design

I. INTRODUCTION

Non-functional properties, such as energy, time, and security (ETS), are increasingly important for the programming of cyber-physical systems (CPS). These systems typically include small embedded devices, which operate under very restrictive ETS constraints, and rely extensively on battery power. Optimising both application execution time and energy consumption while meeting the specified security requirements must be performed early during development to achieve the tight time-to-market windows in this sector with high confidence into the correctness of the CPS. Developing embedded software that respects its specified ETS budget is challenging for developers, who frequently lack tool support to expose and analyse ETS properties during development. Traditional techniques rely on dynamic profiling of code on the embedded device, followed by a time-consuming iterative process of understanding the profiling data and several trial-and-error optimisation attempts. Furthermore, developing software that is optimised for multiple non-functional properties at once is an even greater challenge. The TeamPlay project addressed these challenges by providing the developer with tool support to analyse ETS properties at the source code level together with a methodology for developing embedded applications in which ETS properties are treated as first-class citizens. The TeamPlay approach, therefore, enables developers to reason about the energy, time, and security properties of their applications very early in the process and as an integral part of their typical development activities. The Teamplay methodology comprises a number of key components:

- A cross-layer management of ETS properties from source code to binary, through a complex set of optimisations arising from the different elements of the developer workflow, i.e., the compiler and the scheduler.
- A contract specification language (CSL), allowing ETS properties to be reflected into source code as a first-class citizen, and allowing contracts over those ETS properties.
to be expressed in the source and proven formally using dependent types.

- An explicit coordination layer that takes care of scheduling and mapping decisions on heterogeneous multi-core architectures.

- Comprehensive energy modeling methods for characterising the energy consumption of commercial off-the-shelf (COTS) platforms, and predicting energy usage either statically during compilation or dynamically at runtime.

- Demonstration of the TeamPlay methodology on a number of realistic use cases from the medical, aerospace, autonomous systems, and AI domains where significant improvements in both ETS and in usability have been achieved.

In this paper, (1) we give an overview of the TeamPlay toolchain and methodology for both predictable and complex architectures; (2) we present a high-level overview of the key technical challenges that TeamPlay addressed, including making ETS properties first-class citizens in a transparent manner, enabling the programmer to reason about ETS constraints as part of their development process, energy-modelling targeting multiple different architecture types and generally software development with stringent ETS goals and requirements from the outset; and (3) we demonstrate the TeamPlay methodology on a number of industrial use cases from the domains of medical devices, space, aerial vehicles, and deep learning.

II. OVERVIEW OF THE TEAMPLAY TOOLCHAIN

The TeamPlay approach is focused around providing programmer-friendly abstractions that turn ETS properties into first-class citizens of the language they are embedded in. This allows the programmer to reason about ETS properties as first-class citizens of the language that takes care of scheduling and mapping decisions on heterogeneous multi-core architectures. By exposing ETS properties as first-class citizens and by making them easy to use and understand for application programmers, we enable the programmer to manipulate, and precisely reason about, ETS as normal program values, directly interacting with analytical and optimisation frameworks, including the physical properties of the underlying hardware.

Depending on the criticality and performance constraints of the system, we consider different platforms ranging from simple predictable hardware platforms, e.g., the Nucleo STM32F091RC, to more complex and less predictable ones, e.g., the Apalis TK1. The applicability of some sub-components of the methodology, especially time-oriented ones, strongly depends on the platform choice. In order to enable the TeamPlay methodology on the widest possible range of hardware options we design two specialized workflows, one targeting predictable architectures and one targeting complex architectures as explained in the two subsequent sections.

A. The TeamPlay Methodology for Predictable Architectures

We classify an architecture as predictable if the number of cycles that an instruction takes to execute can be statically determined. The deterministic nature of such processors makes them perfect targets to have their energy consumption modelled at the Instruction Set Architecture (ISA) level, which in turn facilitates analysing security properties by observing time and energy consumption.

Figure 1 depicts the workflow for such architectures. The methodology starts with an annotated version of the application source code in C. The annotations are defined in the Contract Specification Language (CSL) [1]. They capture ETS properties at specific code locations. The CSL layer gathers information about the code structure, e.g., tasks and their parameters, and the associated ETS code points of interests (POIs). The collected POIs are then sent to the multi-criteria optimising compiler (implemented in the WCET-aware C compiler WCC [2–4]), that applies a set of multi-objective optimisations [5] regarding energy usage, timing control, and security risks. To do this, WCC employs three plug-in tools, each of which targets a distinct non-functional property. The aiT tool [6] statically computes a bound for the Worst-Case Execution Time (WCET) for each identified task in the code. The EnergyAnalyser tool [7] uses robust and accurate energy models [8], [9] for the target hardware platforms to predict the energy consumed by each task. The SecurityAnalyser quantifies the security protection level against timing and power side-channel attacks with novel metrics [10], and the SecurityOptimiser transforms the code, if required, to increase protection against these side-channel attacks [11], [12].

In addition to generating object files for the input program, the multi-objective optimising compiler generates a file containing all the ETS properties extracted during compilation. Coupled with the information extracted from the source code, this enables the coordination layer [13] to validate the schedule and generate the necessary glue code for the initialisation, configuration, and runtime management of the tasks [14].

Similarly, the Non-functional Properties Contract System [15], [16] formally proves, using dependent types, that both energy and time budgets as well as the security risk of each identified POI respects the ETS properties extracted by
the compiler. The Contract System generates a certificate that could serve as a proof for certification authorities.

B. The TeamPlay Methodology for Complex Architectures

In contrast to predictable architectures, complex architectures cannot be statically analysed to determine WCETs due to features such as complex pipelines [17] or generally undisclosed architectural and design information. Thus, the static analysers in the workflow for predictable architectures must be replaced by dynamic profilers for complex architectures.

Figure 2 depicts the resulting workflow. In addition to C source code, other programming languages, including C++ and CUDA, can be annotated with ETS properties in CSL, demonstrating the general applicability of CSL annotations. The CSL layer identifies the POIs within the code and extracts the program structure, i.e., tasks and their dependencies. The coordination layer is then executed and, at first pass (solid arrow path), generates glue code to sequentially manage the execution of tasks, while respecting dependencies. The resulting sequential binary is then instrumented with measuring points that, when executed several times, allow us to generate a dynamic profile of the energy consumption and execution time for each task. In this toolflow, energy measurement and execution time extraction are performed using PowProfiler [18], [19]. Reconnecting with the previous workflow, the CSL and coordination layer are provided with the estimated ETS properties of the application. The coordination layer uses this information to schedule the application [20], [21], generating glue code for the management of the tasks accounting for the parallel capability of the underlying platform (dashed path in Figure 2).

III. TEAMPLAY CHALLENGES

A. The Transparency Challenge

During the design of a CPS it is common practice to analyse the ETS effects once the application has been completed. The challenge we address with TeamPlay is to enable programmers to reason about these non-functional properties in an early stage of the development cycle instead. TeamPlay promotes ETS information from the binary to the source code level. Full transparency of ETS properties requires correlating source code features with the corresponding parts of the binary [22]. A contract must be formally proven to guarantee the ETS constraints across the different levels of abstraction. Clear, human-understandable feedback needs to be provided in order to allow the developer to take actions should the application code fail to satisfy some of the constraints.

B. The Energy Modelling Challenge

A fundamental step to enable static energy consumption estimation of embedded software is the ability to accurately characterise and model the energy consumed by the target applications. This requires data collection from each target hardware platform and a model generation methodology. The resulting energy models are directly sourced by static energy analysis tools. Existing techniques may rely on available hardware support or external devices to associate costs with ISA-level constructs. However, these types of models often involve a very deep micro-architectural understanding of the target platform and can require a significant amount of time to develop and validate. Since TeamPlay intends to target different types of architectures, a particular challenge is to design a configurable modelling methodology that captures platform behaviour at a higher abstraction level while yet providing very accurate prediction. Additionally, the methodology must be cost-effective to promote wide-spread adoption.

C. The ETS-aware Development Challenge

ETS properties are frequently considered as unrelated entities that can be accounted for independently of one another, despite the fact that there is substantial evidence that they are sometimes highly correlated. For some architectures, the longer an application executes, the larger the energy consumption and vice versa. However, this relationship is becoming less and less trivial in modern semiconductor systems since leakage power dissipation no longer decreases with reductions in transistor size, so there is often a "sweet-spot" somewhere in the middle of the frequency range where optimal energy consumption can be achieved. This is especially true for modern embedded systems. Similar considerations can be made when security concerns are added, since some security countermeasures may imply executing additional instructions to obfuscate parts of the control flow that would lead to the leakage of sensitive information. Considering all these non-functional properties at once is, therefore, the main challenge addressed in TeamPlay.

WCET analysis is used to determine how much time is spent in the worst case for the execution of a particular piece of software. It is thus a crucial part of the development process of safety-critical embedded systems. Most existing WCET-aware compilation frameworks do not provide a smooth integration of WCET analysis into a compiler [23], or do not scale with industrial-grade applications [24]. Similar considerations apply to the integration of energy consumption and security analysis into compiler infrastructures. In TeamPlay, we targeted full
integration and exploitation of ETS analysers at the compilation level, making the use of these tools within each workflow transparent to the developer.

Worst-case optimisation demands fundamentally different optimisation strategies than average-case optimisation. It is expected that, similar to time, average and worst-case energy usage will exhibit similar fundamental differences. The overall challenge of ETS-aware compilation is thus to identify feasible optimisation approaches in order to balance energy usage, timing, and security guarantees within the compiler, both from the average and the worst-case perspective. Unlike time (e.g. seconds or clock cycles) and energy (e.g. Joules), there is no consensus on one single objective security metric. Existing metrics are usually tied to particular attacks. Consequently, we designed and implemented novel security metrics to counter timing and power side-channel attacks. Here, our focus is on information leakage related to time and energy/power. Schedulers tend to follow the same path as compilers when dealing with energy, time, and security. A challenge in TeamPlay was to raise the awareness of the scheduling strategy to embrace the complete set of ETS properties.

IV. TEAMPLAY USE CASES AND KEY RESULTS

To validate the TeamPlay methodology, we apply it to four industrial-grade use cases. The purpose of the use cases is to evaluate the benefits of the TeamPlay approach in terms of energy saving, timing control, and security improvement. Some purely technical issues that could not be resolved within the project time-frame hindered the evaluation of security on the industrial use cases; thus, our security approach and tools were validated on synthetic benchmarks on the Cortex-M0. All evaluation results are provided in a public deliverable [25].

Through our evaluation, we observed that the TeamPlay methodology gives considerable reduction in development time and effort, and significantly increased confidence into the behaviour of applications regarding their ETS properties and requirements. However, we note that it is typically characteristic that such efficiency and confidence improvements cannot be easily quantified objectively.

A. Camera Pill

For decades, Gastrointestinal (GI) cancer has been one of the leading causes of death in developed countries [26]. Regular screening of high-risk patients with a genetic background can reduce mortality by detecting fatal diseases early. Capsule Endoscopy enabled through camera pills has recently received a great deal of attention as a minimally invasive imaging procedure that offers novel methods for early diagnosis of disorders of the GI tract [27].

A camera pill embeds an imaging system that takes thousands of pictures during its journey inside the GI tract, which are radio-transmitted to an external receiver. The proper functioning of the camera pill is dependent on the timely execution of its software and hardware, as well as very stringent requirements in terms of power consumption, mostly due to its size. Failure to meet timing constraints could result in a misleading or inconclusive diagnosis; failure to meet the energy budget could deplete the battery before reaching the area of interest. Furthermore, the transmitted images are medical data and hence subject to strict privacy regulations. This requires additional security measures to prevent interception.

Technically, the camera pill embeds a single-core Cortex-M0 coupled with a low-power FPGA acting as an image co-processor. During the lifetime of the project, we successfully applied the full TeamPlay toolchain for predictable architectures (Sec II-A), with the exception of the coordination layer, which could not be used due to a need for hardware-specific optimisations in the scheduler. This processor architecture is fully supported by our static analysis tools, enabling safe and accurate time and energy estimation based on highly accurate energy models [9]. We evaluated security requirements and added encryption capabilities as needed. The WCC compiler successfully trades execution time with energy and generated highly optimised code. The coordination layer and CSL both give a green light with a valid schedulability analysis and a certificate proving the specified energy and time budgets being met. Overall, applying the TeamPlay methodology led to an improvement of 18% performance and 19% energy usage over the use of traditional toolchains.

B. Communication in Space Applications

Energy consumption is a major concern in spacecraft, where every little reduction of power consumption that can be achieved without a significant loss of performance is highly valuable. The space-oriented use case includes an image processing and transmission application using SpaceWire [28]. The proper functioning of the full system is dependent on the timely execution of its software and hardware to avoid missing images. Moreover, the harsh environment requires the overall energy consumed to be minimal. Furthermore, the transmitted images can be of critical sensitivity and must be protected from malicious third-party access.

The target hardware device is the LEON3FT-based GR712RC development platform from Cobham-Gaisler [29]. The runtime environment is provided by a Real-Time Operating System (RTOS) named RTEMS (https://www.rtems.org/), which executes the image processing and communication program. Due to its predictable-by-design nature, the LEON3FT is fully supported by the TeamPlay toolchain for predictable architectures. The source code has been annotated for all three ETS properties. A complete energy model has been built [8] and integrated into the EnergyAnalyser tool that is called by WCC for multi-criteria optimisation. The coordination layer and CSL both gave a green light with a valid schedulability analysis and a certificate proving that the code meets its specified energy and time budget. The Coordination layer also generated the required initialisation, management, and configuration code for the application within the RTEMS runtime ecosystem. Applying the TeamPlay methodology led to an energy improvement of 52% while meeting all deadlines.
C. Uncrewed Aerial Vehicle

Uncrewed Aerial Vehicles (UAVs) or drones are increasingly used in a variety of scenarios, enhancing the use of conventional robotic systems with the oversight of an agile and easy-to-deploy platform. TeamPlay had two UAV use cases [30]. The first concerns a search and rescue (SAR) UAV, a drone surveying above the sea seeking lifeboats. Upon detection, a ground base station is notified. The second use case aims at increasing the quality of agricultural produce, relying on diverse sensing mechanisms and facilitating the transition towards precision agriculture (PA), by, e.g., detecting ground hazards. In both cases, fixed-wing drones, i.e., UAVs where propellers provide thrust, wings lift, and maneuvers are performed utilizing control surfaces [31], embed a computing payload connected to a camera [32]. Object detection is performed on the payload, and any results are transmitted to the ground station. The proper functioning of the use cases depend on correct soft-timely execution and low energy consumption. Due to overlapping frames, for instance, missing a few deadlines is possible without missing objects, whereas energy constraints are tighter, directly impacting the flight time and thus coverage area [33].

Three different hardware platforms have been tested: 1) an Apalis TK1; 2) Nvidia TX2; and, 3) Nvidia Nano. In all cases, Linux OS runs use case-specific software, e.g., C++/CUDA computer vision algorithms for detections on the GPU, and is eventually complemented with Robot Operating System (ROS) middleware [31], [33]. We applied the TeamPlay toolchain for unpredictable architectures but, as a result of platform complexity, omitted fact checker and SecurityAnalyser, which both require further investigation.

A study of the SAR use case, available in [34], exhibits an increment in timing control. Using our TeamPlay methodology, we observe an energy improvement of 18%, resulting in the flight time being increased by approximately 4 minutes. Conversely to the SAR use case, the PA use case [31]–[33] utilized merely energy analysis, yet enabled flight time optimisations. When cruising, the mechanical components of the UAV consumed 28 Watts on average, whereas software components consumed between 2 and 11 Watts, with the toolchain enabling in-flight battery-aware schedulability [31].

D. Deep Learning Deployment

The deep learning (DL) use case involves the use of convolutional neural networks (CNNs) for carrying out various image recognition tasks. CNNs are mathematical structures whose implementation is computationally demanding, involving convolution operations between large tensors or multiplication of matrices of large dimensions. The DL use case is a free spot car parking detection application. We assume a camera placed above ground with multiple parking spots on sight and the trained CNN must then determine how many free spots are present and report it to an external receiver. The goal is to study the application of ETS-oriented design which can then be ported to more advanced image recognition systems, such as industrial automation, or the space sector.

Two different hardware platforms have been considered: 1) a single-core Cortex-M0; and, 2) an Nvidia TK1. Our results show that the multi-criteria optimising compiler offers different compiled variants of the same tasks with different energy consumptions and WCET characteristics that can be a great guide for the application designer.

The TK1 version used only the coordination layer of the unpredictable architectures toolchain with a manual extraction of the application structure and a custom energy and time estimation. Results show that the application generated from the TeamPlay toolchain performs similarly as the original human-optimised version both in terms of energy and time.

V. Conclusion

This paper introduced and summarised TeamPlay, a research project supported by the EU Horizon-2020 programme from January 2018 to June 2021.

The goal of TeamPlay was to provide the system designer with a functioning toolchain for developing embedded applications where energy, time and security properties (ETS) are first-class citizens. Our technologies enable developers to incorporate ETS as a design goal of the software, e.g., to meet time or energy budgets while satisfying security requirements. We considered two architectural classes that we label predictable and complex. In this paper we provided an overview of the TeamPlay methodology for both architecture classes. We further demonstrated our methodology on industrial use cases from a variety of sectors, including medical, space, aerial vehicles, and deep learning, where, in some cases, we were able to achieve energy savings of up to 52%.

In the future, we intend to expand our work in a number of directions. The TeamPlay methodology is currently restricted to C and C++ applications for embedded devices; we therefore plan to extend our methodology to cover a much broader range of general-purpose programming languages (such as, e.g., Java, Python or Haskell), as our proposed methodology is, in principle, generic, and the concepts could be ported to other systems. In addition, we intend to expand our energy-modelling techniques to a wider range of hardware types and to improve our methodology and tool support to design software with ETS properties as a major requirement. Preliminary work has shown that program transformation techniques, such as refactoring tool support, would be very applicable here and thus form a natural extension to our methodology.

Acknowledgements

This work was supported by the EU Horizon-2020 project TeamPlay (https://www.teamplay-h2020.eu), grant #779882.

References

