FIT4Green – Energy aware ICT Optimization Policies

Robert Basmadjian, Christian Bunse, Vasiliki Georgiadou, Giovanni Giuliani, Sonja Klingert, Gergö Lovasz and Mikko Majanen

Abstract— Protecting the environment by saving energy and thus reducing carbon dioxide emissions is one of today's hottest and most challenging topics and is of a rapidly growing importance in the computing domain. The motivation and reasons for optimizing energy consumption from ecological and business perspectives are clear. However, the technical realization still is way behind expectations. One reason might be that technical problems range from pure hardware issues (e.g., low-power devices, energy harvesting, etc.) to software to cooling issues. This paper discusses recent findings and first ideas regarding policies and strategies for energy optimization and the development of a generic plug-in for managing data centers, accompanied by the introduction of the concept of "Green Service Level Agreements (GSLA)". We discuss the general structure (generic architecture) of the plug-in and sketch some of the embedded policies. It is also to be noted that all results are part of the recently started FIT4Green project, funded by the European Union.

Index Terms—FIT4Green, Energy, Optimization, Policies, SLA, Data Centre

I. Introduction

Over 500 million host computers, three billion PCs and mobile devices consume over a billion kilowatts of electricity per year. Following predictions of Greenpeace or EUROSTATS ICT consumes an increasing amount of energy, and is estimated to consume up to 20% of the global energy consumption by 2020. Traditionally, systems and network design seeks to minimize network cost and maximize quality of service (QoS). Electrical energy is needed for ICT both to operate and cool the equipment. This insight led to the Green-IT paradigm referring to environmentally sustainable computing or IT. Thus ICT can help reducing energy expenditure by substituting energy-intensive activities (e.g., E-Work, E-Commerce, or E-Learning) to optimizations of the software itself.

Manuscript received May 14, 2010. This work was partially funded by the Commission of the European Union under the 7th Framework Programme, ICT Call 4 – ICT for Energy Efficiency.

Robert Basmadjian and Gergö Lovasz are with the University of Passau, Innstraße 41, D-94032 Passau, Germany (e-mail: {Gergoe.Lovasz, Robert.Basmadjian}@Uni-Passau.de.

Christian Bunse and Sonja Klingert are with the University of Mannheim, A 5, 6, D-68131 Mannheim, Germany (e-mail: {Christian.Bunse, klingert}@uni-mannheim.de).

Vasiliki Georgiadou is with Almende B.V., Westerstraat 50, 3016 DJ Rotterdam, The Netherlands (e-mail: Vicky@almende.org).

Giovanni Giuliani is with HP Italiana srl, via Grandi 4, 20063 Cernusco S/N, Milan, Italy (e-mail: Giuliani@hp.com).

Mikko Majanen is with VTT Technical Research Centre of Finland, Kaitoväylä 1, FI-90571 Oulu, Finland (e-mail: mikko.majanen@vtt.fi).

The price of such substitution processes to the benefit of the environment is partly responsible for the sharp increase in the volume of data centre services. A study of the company Telecomspricing of November 2009 for instance predicts that the data centre revenue across 19 of the EU25 countries will increase with an annual growth rate of 25% per annum between 2010 and 2015 [1] – accompanied of course by a growing impact of data centers on the carbon footprint of mankind

In summary, ICT offers a way forward for reducing the consumption of energy and carbon emission by reducing land and air transport. However, this potential reduction is partially offset by the power used by data centers and computer networks [2]. Network and service providers have electrical costs reaching billions of EUR. Even a fraction of energy savings in networks could lead to reduced financial costs and carbon emissions. The importance of packet networks on energy consumption increases with the "convergence to the Internet Protocol (IP)" paradigm whereby most modes of communication, including mobile telephony, are increasingly supported by underlying packet networks. Since IP networks rely on network nodes and links, the electrical energy used for operating and cooling these equipment, creates a crucial need for research on energy saving strategies for networks. In order to solve these mentioned lacks, recently a great deal of research effort has been dedicated, especially to the following topics:

- Energy-efficient hardware
- Energy-efficient multiprocessor and Grid systems and data centers
- Energy-efficient clusters of servers
- Energy-efficient wireless and wired networks
- Energy-efficient cooling.

In FIT4Green, we aim at the development of a comprehensive view for energy efficiency, involving all layers ranging from technological to business aspects. The focus thereby is on single and federated data-centers supporting different computing styles (traditional-, super- and cloud computing). On a more technical level the FIT4Green strategies and tools includes physical nodes, cooling of nodes, networking hardware, communication protocols, the services themselves that are running on the nodes, up to business plans and service-level agreements (SLAs).

In detail within the FIT4Green project, a set of energy aware scheduling mechanisms and policies will be developed. More specifically, in the case of a single site data center, the idea is to provide algorithms to multiplex, de-multiplex workload in order to save energy. This also includes findings regarding the trade-off between performance, quality-of-service (QoS), and energy consumption, In addition to scheduling issues the improvements of FIT4Green also benefits federated (cloud) environments.

Beneath this optimization the energy consumption might also benefit from software optimizations. Research has shown that especially communication is one of the largest cost factors with respect to energy. This makes communication an ideal candidate for optimizing a system's uptime. Energy optimization has to be aware of the tradeoffs between performance, energy consumption, and QoS. In contrast to hardware optimizations, software systems are usually optimized at development time by specifying their energy characteristics and by adapting the implementation. However, this requires individual adaptations of each system variant, and often implies a negative impact on the performance or QoS of such systems. The FIT4Green challenge is to explore the relations among the various components and to understand the tradeoffs. This enables the development of systems that achieve an optimal balance between performance, QoS, and energy consumption by adapting themselves at runtime (i.e., dynamic optimization).In conclusion, FIT4Green introduces a new paradigm of energy reducing efforts by creating an energy-aware plug-in placed on top of existing data centre automation frameworks. This paradigm will be realized with appropriate models and technology and tools that FIT4Green will develop, implement and test. The remainder of this paper is structured as follows. Section II provides a brief overview on the goals of the FIT4Green project and also gives a short summary of related work. Section III discusses the technical approach for reaching the previously defined goals. Based on this, Section IV introduces the generic plug-in architecture, and finally Section V provides a short summary and conclusions.

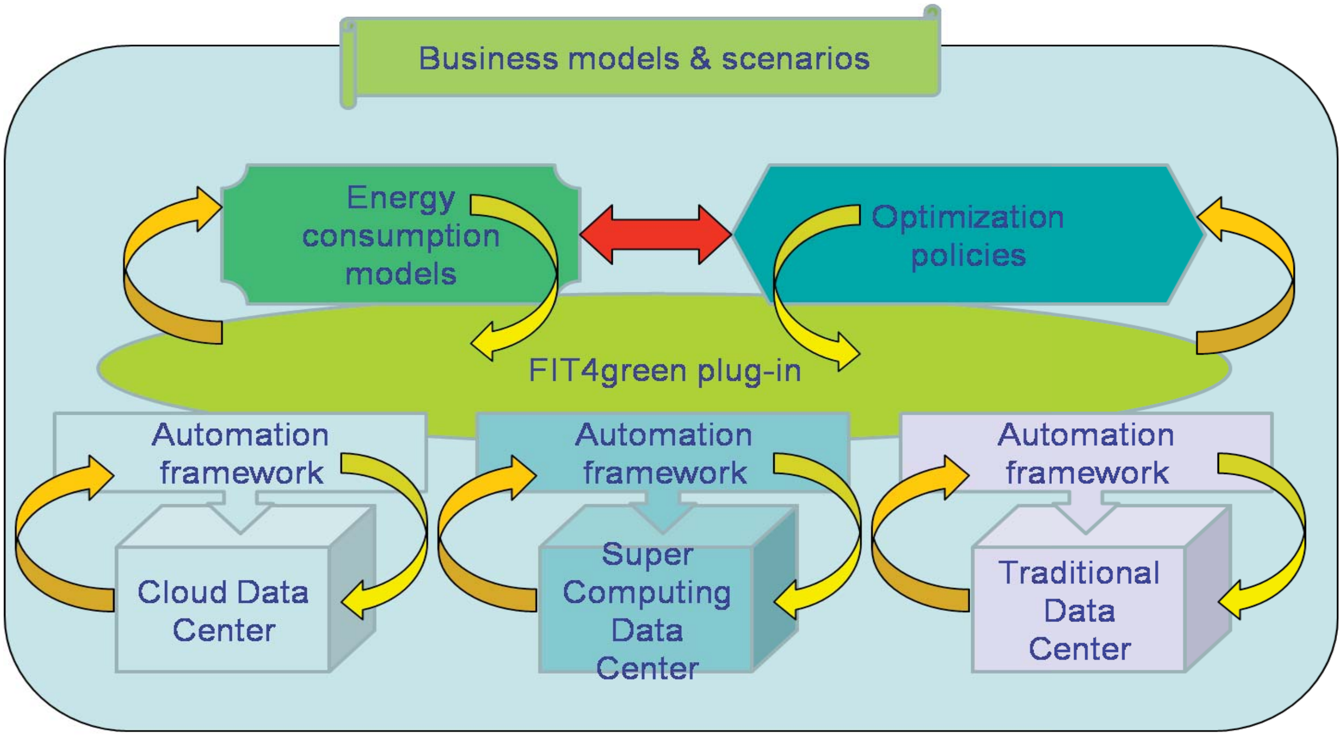

Figure 1. Overview of the FIT4Green technical approach

II. Project Goals

One goal of the FIT4Green project is to develop energy aware optimization policies for data centers, which - once applied - would reduce the energy consumption of their ICT infrastructure, without compromising compliance with Service Level Agreements (SLA) and Quality of Service (QoS) metrics. The FIT4Green approach will be potentially applicable to any type of data centre with any automation framework. Based on the specifics of the data centre and the scenario at hand the percentage of the energy reduction induced by applying these policies can vary; we envision that for a data centre with no previous steps with regard to energy optimization, FIT4Green policies and models can provide on average 20% saving in direct server and network devices' energy consumption and induce an additional 30% saving due to reduced cooling needs. The first estimate is based on recent "Data Center Energy Forecast Report" in which Accenture [3] illustrated a "State of the Art" exploitation of consolidation and optimization strategy of IT computing resources inside a data centre in Silicon Valley. Additional savings are possible via the reduction of cooling energy. This can be concluded from a study by HP and the Uptime Institute [4] that shows that "most data centre power is spent on cooling IT equipment (between 60 and 70%)".

Additionally, the federation of data centres will allow lowering the overall Green House Gas (GHG) emissions (e.g., CO_2 emissions) by relocating applications and services to sites where lower environmental-impact power-generation may be available.

In addition to the technical goal to prove the energy saving potential of FIT4Green policies, the consortium aims at creating a lasting impact by supporting the data centre industry in the adoption of the FIT4Green policies instantiated in the plug-in. One issue in this context is the design of Green-SLAs that account for the modifications of the service delivery induced by FIT4Green. Another issue is to acknowledge the fact that a positive evaluation of the economic ROI is the number "1" criterion also for environmentally targeted investment decisions. Both issues will have a bearing on the design and especially on the exploitation phase of the project.

III. Technical Approach

Lately, many ICT players have been proposing focused solutions to optimize single component's energy consumption (low consumption servers, CPU speed scaling, power save modes, efficient cooling devices for data centres, etc.). These solutions enable the energy footprint reduction of single devices, treating them as isolated components and therefore lacking any savings obtainable through a holistic approach. Consolidation and virtualization techniques lead to energy savings through the reduction of the number of active servers; however, the current deployment strategies are fairly static and not guided by energy saving principles. Also, current SLAs do not include any metrics related to the ecological footprint.

FIT4Green goes beyond this state of the art with the global analysis of IT solutions deployment needs and the optimized deployment scheme inside a single site data centre as well as a federation of data centres with different energy related characteristics, considered as a global resource pool. By (re-) distributing computation and resources among several data centres, FIT4Green is able to capitalize on additional degrees of freedom with potential high impacts on the optimization strategies at federation level using, for example:

- Data centres at different geographical latitudes with the respective implications on cooling requirements based on external temperature, possibility to recycle heat through co-generation devices (different latitudes in the same hemisphere – capitalizing on temperature range variations – or different hemisphere – for seasonal changes)
- Data centres in regions where different sources of power generation are available, at different costs and GHG emissions
- Data centres inside different time zones with respect to the clients, allowing a different balance or mix of computing tasks.

Taking a global view on IT solutions (rather than a focused view on single components) and applying global optimization throughout the whole ICT-based process, FIT4Green technical approach (see Figure 1) includes the following topics:

- An optimization layer on top of existing data centre automation frameworks – integrated as a modular "plug-in" – to guide the allocation decisions based on the optimized energy model-based policies.
- Energy consumption models will be developed, and validated with real cases, for all ICT components in an IT solution chain, including the effects due to hosting data centres in sites with particular energy related characteristics, like alternative power availability and energy waste/recycle options, etc.
- Optimizations, based on policy modelling descriptions able to capture the variety of deployment that are possible for a given application or a set of applications, integrated with specific attributes supporting the evaluation of energy consumption models, will guide the deployments decisions on which/when equipments need to stay on, where/when applications should be deployed/relocated, also capitalizing on the intrinsic non linear behaviours of energy consumption growth with respect to the load of ICT components.

SLAs and business models will be analysed with respect to their potential impact on the carbon footprint of ICT. Consolidating these findings with implications of a deployment of the FIT4Green policies, new Green-SLAs and Green business model components will be developed taking into account the carbon footprint of data centre services. These will be integrated into the FIT4Green exploitation strategy.

Obviously, by moving applications and services to alternate data centres force the network traffic between client and servers to follow different paths, which have different impacts on the global energy consumption of the full process: both networks operated by telecommunication operators and local area networks will be considered in the global optimization schemes.

In this context, the typical distribution of clients is very significant as well: services with a global scope, e.g. search engines or Web 2.0 applications, receive requests from all over the world (following some distribution patterns bound to local times); on the other hand a local public administration service provider will most likely receive requests from a much more limited geographical area (and therefore time of day interval). The effect of the deployment of such services inside a federation of data centres implies a completely different parameterization of the model for the computation of the energy consumption: long term relocation of the public administration services to a data centre in the opposite hemisphere will force all users to have a much longer network path, while such effects do not show up so strongly if the services has an almost homogeneous distribution of clients around the globe. The picture changes again if it is possible to quickly relocate services, for instance with a "follow the sun" pattern; the energy impact of the service relocation (transition) needs to be considered in the overall computation for the optimal solution.

In the first phase of the project a set of FIT4Green scenarios was developed: The computing styles traditional, cloud and supercomputing are dealt within the context of single site and federated site data centres. Each scenario highlights the special feature of the particular setting: The traditional computing scenarios, for instance, mainly deal with the challenge of deploying the FIT4Green energy saving strategies under the constraints of the data centre automation framework reacting rather slowly to the FIT4Green policies suggested by the plug-in. The automation framework in the cloud computing scenarios is much more flexible; however the plug-in has to cope with an unknown variety of applications and unforeseeable spikes in demand. There will be one pilot site for each computing style. Service/Enterprise Portal, Grid and Cloud pilots will support both single site and federated sites scenarios: multiple collaborating data centres inside the same organization for the Portal pilot; federation of supercomputer systems for Grid and open cloud federation of multiple labs for the Cloud.

Each pilot will implement the respective scenarios, measure the overall energy consumption and related cost reduction, apply the optimizations, evaluate the results and assess the expected impacts; this process will be iterated three times to allow energy models and optimizations to be refined based on the results collected from the real test beds.

Finally, FIT4Green will investigate on the lessons learnt in the development of energy models, optimization policies and plug-ins for the various computing styles, and rationalize them in a set of guidelines for the development of future IT solutions, which will intrinsically consider energy consumption and environmental footprint as essential design principles.

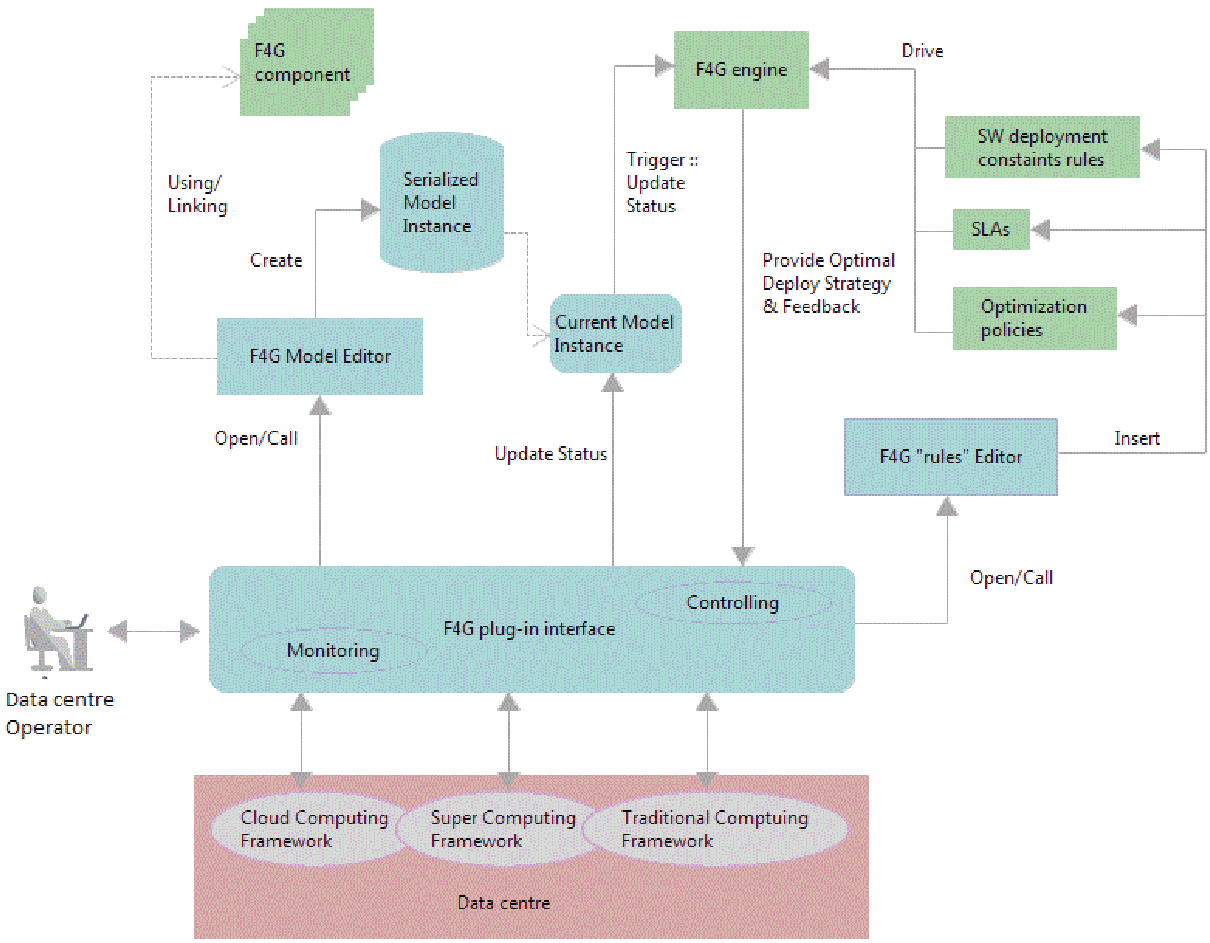

Figure 2. Schematic overview of the FIT4Green general architecture

IV. General Architecture

Traditionally, the development of plug-ins for management & control software of data centers has followed the semantics defined by control design approach. In this context maintainability, fitness-for-change, coupling and cohesion are considered to be important structuring criterions. It is easier to keep consistency and completeness of information with a structured design approach. In the context of the FIT4Green project coupling and cohesion criteria as well as a model-based development process (e.g., using UML) have been considered for mapping functional requirements into components.

Moreover, for domain systems (i.e., the data center domain), a reference-architecture is needed that represents a domain specific way of structuring the control software plug-in through decomposing the problems into parts and their relationships, and mapping them to software units and their interactions. To systematically achieve this structure of the plug-in, its functional and non-functional requirements as well as architecture styles and patterns are needed. Architecture styles denote well-known ways of structuring. Thus, the FIT4Green project follows an architecture-based development process comprised by the following steps:

1. Developing subsystems for the requirements: A set of subsystems is generated from functional and non-functional requirements, based on architectural styles and patterns.
2. Determining an actual architecture: These subsystems can be seen as components in a larger subsystem. Thus, functional view is described and transformed into a process view based on the considerations of parallelism.
3. Validating the solution: The architecture solution is validated using the quality scenarios, e.g., change scenario for modifiability, use scenario for performance, etc.

In the context of this process the systematic, concise and precise description of the plug-in structure(s) is of uttermost importance. It is the basis for all design activities including comprehending, communicating, analyzing, trading-off, as well as for modifications, maintenance, and reuse.

In the context of FIT4Green the architecture specification is based on mathematical, textual, and graphical notations. In order to manage the plug-in's inherent complexity the overall, general architecture specification is divided into multiple views.

Figure 2 presents the schematic overview of the FIT4Green general architecture. The FIT4Green plug-in is placed on the top of the existing data centre automation and management frameworks. The plug-in is divided into two parts: monitoring and controlling. The monitoring part updates the dynamic parameters of the current meta-model instance that describes the status of the data centre under optimization. The meta-model is initially built up by the data centre operator through the FIT4Green model editor. The latter offers to the operator a variety of FIT4Green components which may be used and linked in order to provide a model of both the structure and the features of the data centre. The FIT4Green components model various ICT components and their energy consumption.

The controlling part drives the FIT4Green optimization engine that finds the optimal deployment actions to reduce the energy consumption of the data centre with regard to the current status, SLAs and rules set up by the FIT4Green plug-in user. The rules consist of SW constraints and FIT4Green policies. Both constraints and policies can be edited with editors. The optimal deployment actions are reported back to the FIT4Green plug-in which invokes the appropriate data centre framework to execute them.

V. Summary and Conclusion

Recognizing that human-made greenhouse gas emissions are the major reason for global warming (i.e., green-house effect) created the urgent need to tackle environmental issues by adopting environmentally sound practices. This leads quite naturally to environmentally sustainable computing or IT, especially in the (large) data-center domain.

Within this paper the FIT4Green approach energy-aware computing for single or federated data-centers following different computing paradigms was introduced. FIT4Green provides energy saving strategies and policies and package these into a plug-in that can be used in the context of data-center control frameworks. Within the paper a generic plug-in architecture was introduced that outlines the general structure and approach. It is important to note that the architecture is kept as generic as possible in order to allow for "easy" instantiation and porting.

Based on individual assessments of the FIT4Green application partners it is currently expected to reach direct savings of 10-30% of the energy costs of a site. In addition, one can expect to save additional energy by reduced cooling needs. As soon as the FIT4Green policies are formally specified and implemented these hypotheses will be evaluated by several industrial-scale case studies.

Acknowledgment

The authors would like to thank all the partners and colleagues working in the FIT4Green project (http://www.fit4green.eu/).

References

[1] http://www.telecomspricing.com/news_detail.cfm?item=2592

[2] X. Fan, W.-D. Weber, and L. A. Barroso. "Power provisioning for a warehouse-sized computer", ISCA '07: Proc. 34th Annual International Symp. on Computer Architecture, pp. 13-23, 2007.

[3] "Data Center Energy Forecast Report" by Accenture: Final Report, July 29, 2008.

[4] C. Malone and C. Belady. "Metrics to Characterise Data Centre & IT Equipment Energy Use", Digital Power Forum, Richardson, Sep 2006 & C. Belady, "How to Minimise Data Centre Utility Bills", Sep 2006.