Seventh FRAMEWORK PROGRAMME
FP7-ICT-2007-2 - ICT-2007-1.6
New Paradigms and Experimental Facilities

SPECIFIC TARGETED RESEARCH OR INNOVATION PROJECT

Deliverable DF1
"Links between Research and Experimentation"

Project description
Project acronym: **ECODE** Project full title: **Experimental Cognitive Distributed Engine** Grant Agreement no.: **223936**
Document Properties
Number: FP7-ICT-2007-2-1.6-223936-DF1 Title: Link between Research and Experimentation Responsible: Dimitri Papadimitriou (ALB) Editor(s): Dimitri Papadimitriou (ALB) Dissemination level: Public (PU) Date of preparation: **December 13, 2010 (v1.1)** **August 31, 2010 (v1.0)**

DF1 - Links between Research and Experimentation

Executive Summary

This deliverable describes the links between research and experimentation by i) proposing a systematic experimental methodology, ii) identifying the tools part of the experimental research chain, iii) by defining the relevant criteria and metrics that shall be verified by results produced by means of experimentally driven research. Basically, this document tries to determine the methodology that systematic experimental research (and its outcomes) shall follow as well as the sine qua non conditions and properties that its results shall meet so as to position experimental research as part of the research continuum from analytical/formal models/theory to experimentation with real-systems.

The results of our investigation led to a presentation entitled "*Experimentation as a research methodology to achieve concrete results: where, how, when*" together with the definitions of the criteria and metrics relevant to experimentation, provided by the ECODE project during the Future Internet Assembly (FIA). This has been held in Stockholm, Sweden on November 2009. This presentation material was accompanied by a two-page document entitled "*On the existence of experimentally-driven research methodology*" which argues that if the objective is to position experimental-research as the corner-stone to the so-called Future Internet research, then the results it produces shall follow identical rules and constraints that other experimental sciences fulfill both in terms of experiments realization but also in terms of results production, verification, and validation. For this purpose, a set of relevant criteria and metrics that experimental results shall meet have defined. The content of this document led to the "*Experimentally driven research*" white paper published by the FIREWorks support action on April 2010. This deliverable provides an extended version of this initial document.

List of authors

Affiliation	Author
ALB	Dimitri Papadimitriou
IBBT	Brecht Vermeulen
CNRS	Philippe Owezarski

Table of Contents

1. Introduction

FIRE aims to create a “research environment for investigating and experimentally validating highly innovative and revolutionary ideas” towards new paradigms for future internet architecture by bridging multi-disciplinary long-term research and experimentally-driven large-scale validation. FIRE foundational objectives were:

- Creation of a multi-disciplinary, long term research environment for investigating and experimentally validating highly innovative and revolutionary ideas for new networking architectures and service paradigms;
- Promotion of experimentally-driven yet long-term research, joining the two ends of academy-driven visionary research and industry-driven testing and experimentation, in a truly multi-disciplinary and innovative approach;
- Realization of a large scale European experimental facility, by gradually inter-connecting and federating existing and new “resource clusters” for emerging or future internet architectures and technologies.

These objectives further evolved toward the inception of experimentally-driven research as a visionary multidisciplinary investigation activity, defining the challenges for and taking advantage of experimental facilities. Such investigation activity would be realized by means of iterative cycles of research, oriented towards the design and large-scale experimentation of new and innovative paradigms for the Internet - modeled as a complex distributed system. The refinement of the research directions should be strongly influenced by the data and observations obtained from experiments performed at previous iterations thus, being “measurement-based” which in turn requires the specification of the relevant criteria and metrics as well as their corresponding measurement tools.

The rationale was thus clear: create a dynamic between elaboration, realization, and validation by means of iterative cycles of experimentation. Its realization was already less obvious and rapidly confronted to the lack of computer communication/networking experimental model. Moreover, the “validation by experimentation” objective opens a broad spectrum of experimentation tools ranging from simulation to real system experimentation. The selection of the experimental tools depends on 1) the object of experimentation (corpus), 2) the nature and properties of the results, and 3) the cost function that itself depends on complexity, experimental and running conditions but also on the level of abstraction (referred to as “realism”).

Our thesis is that experimental validation of “elaboration and realization” requires a broader set of tools: starting from more abstract tools (not only because their resulting cost is lesser but because such tools produces results verifying all conditions explained here below) followed by progressive addition of realism as part of the experimented system to ultimately reach the so-called field trials with real systems. Thus, systematic experimentation is a continuum. The following sections describe the dependencies with respect to this experimental chain and its associated set of criteria and metrics.

2. Systematic Experimental Methodology

Computer communication/networking is characterized by two fundamental dimensions: i) distribution of a large number of dynamically interacting (non-atomic) components, and ii) the temporal variation of their inner properties that in turn influence these interactions. A couple of examples would set the landscape: autonomic networking is the transposition of the autonomic computing concept in the communication space, and network “virtualization” is the transposition of the abstraction concept of object-oriented programming in the networking space. More, the dynamic nature of these interactions results in modifying its scaling properties of the individual components besides modifying the properties of the global system. Many other examples can be cited, the fundamental observation is: no experimental model actually exists – or more precisely – the complexity of the resulting system makes its modeling a research discipline on its own.

This doesn’t mean or imply however that an experimental methodology could not be defined based on i) a broader set of tools ranging from simulation[1] to real system experimentation and ii) our experience from practicing core experiments in the computer communication/networking disciplines. Such methodology would include the following steps (part of each iteration):

i) Specification of the performance objectives, (technical and non-technical) constraints, and description of expected results

ii) Definition of relevant performance criteria and metrics

iii) Description of the modus operandi including configuration, initialization, running conditions, and (iterative) procedure(s) to be executed

iv) The reporting on observations and the resulting analysis as well as the feedback after performing each iteration before reaching (partial) conclusion.

[1] To keep this document simple we do not distinguish between the various classes of simulation from model simulation (macroscopic) to procedural simulation (microscopic) nor distinguish between the various class of simulation techniques

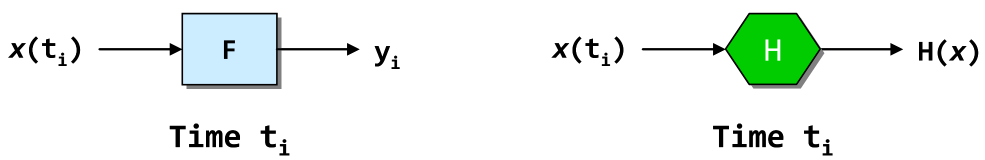
3. Experimental Criteria and Metrics

In order to ensure verifiability, reliability, repeatability, and reproducibility of the experimental results produced, one shall characterize the output of experimentation. Meeting these criteria implies in turn to control the parametrization, the input and output, as well as the running conditions of the conducted experiments. On one hand, verifying the repeatability, reproducibility, and reliability criteria enables generalization of the experimental results produced. On the other hand, ensuring verifiability of these results increases their credibility (results can be "explained").

Let's now proceed with the definitions of the criteria that experimental results shall verify. If we define a given experimental model by a function F, with input variables $x_1,...,x_n$ and parameters $e_1...,e_m$ such that $F(x_1,...,x_n|e_1,...,e_m) = y$, then the following properties hold:

3.1 Verifiability

Verifiability means that we can find a formal model H of F such that $H(x_1,...,x_n|e_1,...,e_m) = F(x_1,...,x_n|e_1,...,e_m)$.



$$\exists\ H: \mathfrak{R}^n \rightarrow \mathfrak{R}: x \rightarrow H(x)$$
$$\text{such that } \mathbf{H(\mathit{x}(t_i)) = y_i}$$

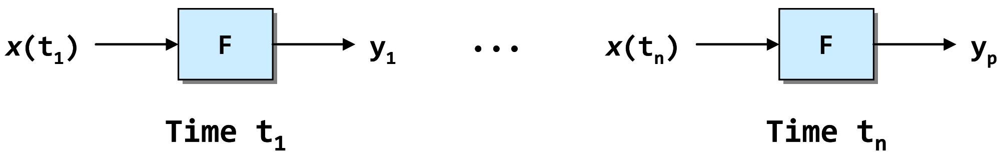
It is important to underline the distinction between verification (verifiability) and validation (validity). Verification means that the experimental model output should satisfy the formal model output (e.g. computational model). On the other hand, validity is formally defined as follows: a proposition A is valid if H(A) = TRUE for any model H of A; thus, in experimental research, we can only verify satisfiability (A has at least one model H for which H(A) = TRUE). Note that in propositional logic, one usually verifies validity by applying the following theorem: a proposition A is valid if and only if its opposite (negation) can not be satisfied. Henceforth, the best we can hope concerning "verification" is to find at least one model H of A that verifies the same "output" as the realization F of A by experiment: H(A) = F(A). If this is the case, then one does indeed satisfy the initial proposition (it is verified by one model) but not validate it (the proposition is not verified for any model).

Also, one constructs (independently) a model to verify that the output of the experiment F satisfies to the output of the model H. Thus, one does not verify the conformance of experimental execution against the specification of the experimented system but the consistency of the experimental output against a computational model of the experiment drawn independently from it. Verifiability is thus not synonym of conformity test or conformance test against the specification of the underlying experimented system.

Finally, it is interesting to observe from its definition that verification is the formal complement to experimentation (instead of positioning experimentation as the complement to the theoretical model).

3.2 Reliability

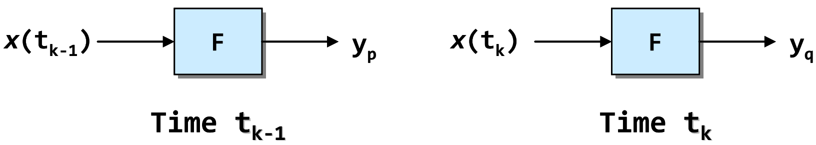
Reliability (defined as the probability that system or component will perform its intended function during a specified period of time under stated conditions) means that output of the model during the pre-defined time interval $[t_{k-1},t_k]$, $1 \leq k \leq T$, $F(x_1,\ldots,x_n|e_1,\ldots,e_m)[t_k] = y[t_k]$ exists within a pre-defined range of valid output.



$$\exists\ [t_1,t_n] \text{ such that } \forall\ k \in N,\ 1 \leq k \leq T$$
$$F(x(t_k)) = y \text{ exists and } y \in [y_1,y_p]$$

3.3 Repeatability

Repeatability means that if $(x_1,\ldots,x_n|e_1,\ldots,e_m)[t_{k-1}] = (x_1,\ldots,x_n|e_1,\ldots,e_m)[t_k]$ such that $y[t_{k-1}] = y_p$, $y[t_k] = y_q$ then $y_p = y_q$ (reliability is thus characterized by persistence of the output in time).



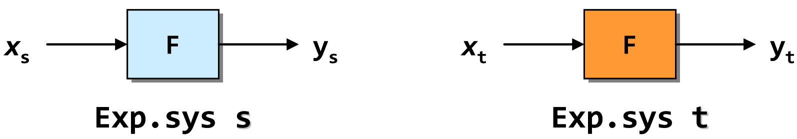
$\forall\ k \in N$

if $x(t_k) = x(t_{k-1})$

then $y_q = y_p$

3.4 Reproducibility

Reproducibility means that the experimental model $F(x_1,\ldots,x_n|e_1,\ldots,e_m)$ can be executed at the same time (simultaneously) on different experimental systems and produce the same output if the input to both models is the same (reproducibility is thus characterized by persistence of the output in space) .



$\exists\ s,\ t \in S$

if $x_s = x_t$

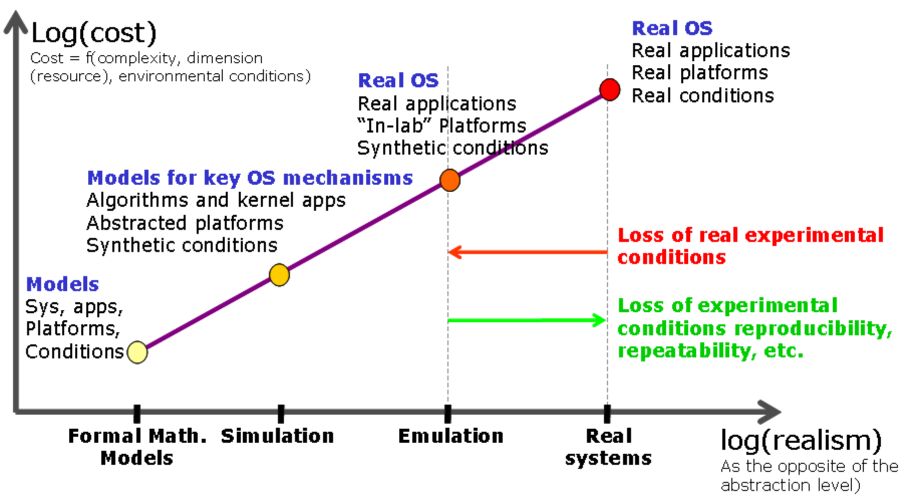
then $y_s = y_t$

4. Experimental Tools

Different experimental tools can be used. As stated above their selection is neither arbitrary nor religious: it depends on the experimental objective and maturity of the experimented corpus. Nevertheless, each of them needs to ensure that the conditions defined here above are verified.

However, it is clear that fulfilling these conditions does not come at the same cost for the same level of abstraction. We can distinguish three types of abstraction: i) abstraction of the network/shared infrastructure (network resource consumption model, processing model, etc.), ii) abstraction of the system (processing/memory resource consumption model, computation model, etc.), and iii) abstraction of the environment (traffic model, application model, user/behavior model, etc.). To each (non-atomic) element of this partition of the abstraction space, we can associate a level of realism when the abstraction is replaced by a “real” entity. Without entering into the debate of reality or what reality actually represents or means, we simply consider here a real system as an instantiation of the experimented component models at the hardware and/ or software substrate level depending on the expected level of performance.

In this context, validation of a new routing algorithm, for instance, would be better conducted on a simulation platform (after formal verification) not only because their resulting cost is lesser but because such tools produces results verifying all conditions explained here above. Afterwards, progressive addition of realism as part of the experimented system would consist in instantiating the execution stratum (remove the system abstraction) in order to perform emulation experiments. Emulation experiments can lead to reproducible and repeatable results but only if their "conditions" and their "executions" can be controlled. Realism can thus be improved compared to simulation (in particular for time-controlled executions of protocol components on real operation system). Nevertheless, such experiments are more complex and time consuming to configure and execute; performance evaluation is possible if the experimental platform comprises a “sufficient number” of machines (representative of the experiment to conduct). Emulation still requires synthetic network conditions (models) if executed in controlled environment and either injecting real traffic or replay traffic traces (not that even when available "spatial distribution" of traffic is available remains problematic to emulate because the spatial aggregation of address prefixes is not necessarily as the routing tables are often not provided together with traffic traces).



Stepping into real system experimentation increases cost but increases realism. As such the loss of control of experimental conditions in such systems raises the issue of persistence of the properties observed earlier in the experimental chain. In particular, these properties shall already be determined by the earlier experimental stages (leaving them intrinsically part of experimental research activities).

Practically, in order to ensure - at least better control of the experimental conditions, the following elements might be considered:

a. Specify performance analysis methodology together with the necessary mathematical tools to be able to perform data analysis and mining tasks on experimental data coming from various monitoring points (from single or multiple testbeds). This objective also covers specification the necessary mathematical tools to analyze the sensitivity of the performance measures to changes in the "experimental model" parameters. Sensitivity analysis attempts to identify how responsive the results of an experimental model are to changes in its parameters: it is an important tool for achieving confidence in experimentation and making its results credible. Sensitivity analysis quantifies the dependence of system behavior on the parameters that affect the modeled process and in particular its dynamics. It is used to determine how sensitive a model is to changes in the numerical value of the model parameters and changes in the model structure--
b. Specify distributed performance monitoring system (while) allowing experimenters to choose the best tool(s) for their experimentation.
c. Define a standard experiment description and control interface and wrap existing tools within this API. This standard interface will focus on providing a common programming interface to describe every aspect of a networking experiment but will also attempt to provide robust experiment monitoring and management facilities and will integrate with the data analysis and data mining tools developed in a). Note: sensitivity analysis of the reliability, the performance, and the performability of monitoring system is a complementary objective.

Note: Sizing experimental environment vs scalability experiments

The scale of a system is measured by the rate x state x size that the system can sustain when running using a given number of resource units (for processing and storage). Networking systems can thus only scale indefinitely if and only if the rate of change of the state set, the number of states and the size of each state are independent of the global properties of the environment into which the system is operating. It is thus fundamental to mention that the scale of an experimental facility (the number of resource units and their distribution) does not determine the scalability properties of the corpus. However, the scalability properties of the experimented corpus determine the number of resource units that are locally required to be executed at a certain scale. Thus, such experiment can be performed to i) verify a pre-estimated level of scaling of the experimented corpus and/or ii) iteratively determine the scale property of the corpus with the risk that the dependency on the global properties could never be found (hidden dependencies, correlations, non-linearities, etc.). Hence, only the former leads to verifiable experiments. In other words, a large-scale facility can only help verifying scaling properties but not determine these properties. Positioning and role of so-called large-scale testbeds is further discussed in Deliverable DF2.

5. Conclusion

Starting from the initial objectives of the FIRE initiative and its associated methodology (dynamic between elaboration, realization, and validation by means of iterative cycles of experimentation), this paper argues that a broader set of experimental tools is required to implement systematically this methodology. Indeed, this paper positions systematic experimentation as a continuum: starting from more abstract tools (not only because their resulting cost is lesser but because such tools produces results verifying all criteria explained in Section 3) followed by progressive addition of realism as part of the experimented system to ultimately reach the so-called field trials with real systems. The addition of realism at increasing cost (resulting from the increasing complexity) is the main purpose of performing experimentation by means of emulation or real systems. In other terms, systematic experimentation can not be limited to trials on emulated platforms to achieve verifiable, reliable, repeatable, and reproducible results at best cost-complexity (thus experimentation time). Indeed, emulation experiments can lead to reproducible and repeatable results but only if their "conditions" and their "executions" can be controlled. Realism can thus be improved compared to simulation (in particular for time-controlled executions of protocol components on real operation system). Nevertheless, such experiments are more complex and time consuming to configure and execute.