



B E T S Y

Deliverable D4e
BETSY Framework Evaluation and
model validation report

Public



Project Number	:	IST-004042
Project Title	:	BETSY
Deliverable Type	:	Report

Deliverable Number	:	D4e
Title of Deliverable	:	BETSY Framework Evaluation and model validation report
Nature of Deliverable	:	Report, Public
Internal Document Number	:	betsy-phr-d4e-v1.doc
Contractual Delivery Date	:	31 March 2007
Actual Delivery Date	:	15 May 2007
Contributing WPs	:	WP6, WP4, WP3, WP2, WP1
Author(s)	:	Peter van der Stok (Philips)

Abstract

This document describes the evaluation of the work-package results developed during the BETSY project.

Keyword list

Evaluation results

Table of Contents

Table of Contents.....	2
1 Introduction.....	3
2 Assessment approach.....	4
3 Evaluation criteria	5
3.1 End-to-end QoS chain partitioning.....	5
3.2 Limited capacity of resources.....	5
3.3 Energy consumption.....	6
4 Tangible results	7
5 Assessment of the BETSY objectives	9
5.1 Task 2.1 Identification of end-to-end delivery chain parameters	9
5.2 Task 2.2 Stream description methods.....	10
5.3 Task 2.3 Decomposition of end-to-end constraints.....	11
5.4 Task 3.1 Parameter and resource requirements.....	12
5.5 Task 3.2 Device functional component modeling	12
5.6 Task 3.3 Network functional component modeling.....	13
5.7 Task 3.4 Model composition	13
5.8 Task 4.1 Specification of the BETSY Framework.....	14
5.9 Task 4.2 Design of the test bed and the demonstrator	15
5.10 Task 4.3 Reference implementation of the BETSY framework on the test-bed.....	16
5.11 Task 4.4 Porting of reference implementation on the demonstrator set-up	16
5.12 Task 4.5 Validation and evaluation of BETSY models	17
6 Conclusions	18
References	19

1 Introduction

This document describes the assessment of the results of the BETSY project [1] measured at the end of the project. The quantity and quality of the results of the BETSY project at the task level are measured against the initial objectives.

The deliverable D1d describes the evaluation criteria on the appropriateness of the chosen and developed BETSY techniques. It provides a first insight in the actual technical objectives of the BETSY project. As remarked by the reviewers, these technical objectives should be visible in the demonstrator. This document also takes into account the criteria of success for the final demonstrators as the consortium saw them in the *Evaluation Plan* (D6b) and the *Risk Management Plan* (D6c).

A staged assessment approach was proposed that is based on the individual tasks of the BETSY project. In the next section the proposed assessment approach is presented after which the results of the approach are described.

The described tasks are directly related with the technical objectives. That means that work-packages wp5 and wp6 are not considered in this document. Work-package 1 has prepared the terrain for wp2-wp5 and does not contribute directly to the BETSY objectives. Therefore, wp1 is not considered either.

2 Assessment approach

For each task a table is inserted in which figures:

- The name of the task
- The objectives of the task taken over from the *Description of Work (DOW)*
- The success criteria which determine whether the objective is successfully met
- The impact of failing to meet/ partially meeting the objectives.

Accordingly, the success of the BETSY tasks is evaluated. The success of a given BETSY task can be expressed as follows:

- When the task has reached its objectives, the task is completely successful
- The objectives have not been reached but the fall-back scenario has been implemented. The task is considered partly successful.
- The objectives have not been reached and none of the fall-back scenarios were realized. In that case the task is considered failed.

The evaluation process is targeted to the individual tasks. The overall project success depends on the sum of all accesses and failures. Contributions of the success of the results of wp2 and wp3 are as important as wp4. The success of the test-bed and prototype should not be equated to the technical success of the BETSY project. A rough indication might be that wp2 and wp3 cover 70% of the technical success of BETSY and wp4 the remaining 30%.

This evaluation only tells something about the technical success criteria according to the DOW. For the final evaluation other factors come into play like the timeliness of the delivery of the deliverables, the dissemination success, the contribution to standards and the impact on the scientific and industrial communities. All these aspects are not covered by this document.

This document builds on the deliverable D1d *Evaluation Criteria*. This document looks at the individual objectives of the tasks without taking into account their mutual relations. This deliverable also presents the higher-level success criteria of the technology presented in D1d.

3 Evaluation criteria

The scenarios of D1a show how users perceive energy consumption and resource management. The scenarios show that they are relevant subjects to effectively deploy the video streaming devices. A set of techniques is needed to effectively manage the resources and save energy. The BETSY DOW describes the techniques that are envisaged. Of course other techniques may pop up in a later stage. Under such an occurrence this document needs to be adapted.

In this section the evaluation criteria are derived from the scenario analysis. The criteria can be classified in three classes according to the Description of Work (DOW) [2]:

1. End-to-end QoS chain partitioning
2. Limited capacity of resources
3. Energy consumption

Point 1 is the subject of WP2, while point 2 is the subject of WP3. Point 3 should emerge naturally from both WP2 and WP3.

3.1 End-to-end QoS chain partitioning

Criterion1. A set of stream parameters can sufficiently describe the display quality.

Criterion2. For the functions selected and specified for the BETSY project, mathematical functions describe the effect of a given BETSY function on the stream parameters.

Criterion3. It is possible to express for a chain of functions, the effect on the stream parameters as a composition of the mathematical functions, which describe the BETSY functions.

Criterion4. It is possible to formulate perceived display quality optimisation criteria to optimise the chain of functions with respect to resource usage, delay, display quality, and video quality.

Criterion5. It is possible to optimise a given chain of functions with respect to perceived video quality and the display quality.

3.2 Limited capacity of resources

Criterion6. It is possible to define criteria, which determine whether a new stream can be accepted with the specified quality. The transport of a refused stream leads to unwanted quality degradation in the transported stream. An accepted stream can be streamed and displayed with the specified quality.

Criterion7. It is possible to determine at which quality level a stream can be accepted with the given resource occupation by other streams.

Criterion8. It is possible to determine at which quality level a stream can be continued while either the source or the destination is mobile. The quality level is calculated when a stream needs to change resource usage during movement (e.g. hand-over from one AP to another, different path over network, etc.)

Criterion9. It is possible to calculate the new resource consumption for an existing stream after a quality change of the video

Criterion10. It is possible to relate codec settings to display qualities, such that the display quality can be optimised by changing the codec settings.

Criterion11. It is possible to estimate resource consumption fluctuations under bandwidth fluctuations.

Criterion12. It is possible to embed the models into systems, which have to use also proprietary components (e.g. OS, NIC drivers, etc.).

3.3 Energy consumption

Criterion13. Criteria exist to decide energy consumption settings.

Criterion14. Relations between energy consumption and display quality can be quantified.

Criterion15. It is possible to determine power consumption costs of different mappings between the BETSY functions and the resources.

Criterion16. It is possible to estimate the impact of changes in the mapping or in the parameter values of the BETSY functions on the battery lifetime.

Criterion17. It is possible to correlate the power or energy costs of a number of functions with the power or energy cost of an application/steam composed of these functions.

Criterion18. It is possible to extend the lifetime of the battery by the usage of the information embedded in the models.

Criterion19. When estimations are used, it is possible to determine the degree of confidence in the accuracy of estimations.

Criteria 5-8, 10, 12, 15-16 should also apply on the reference implementation as envisaged in WP4.

4 Tangible results

The validation criteria say something about the technical and scientific soundness of the BETSY project results. The results themselves need to be disseminated and possibly utilized. From a utilization point of view, tangible results coming from the BETSY project are the first step. The criteria of section 3 are equally applied to the software and theory underlying these tangible results in the evaluation of section 5.

There are six tangible results coming from the BETSY project: (1) MPEG-4 encoder, (2) BETSY FrameWork, (3) power consumption models, (4) video streaming delay calculations, (5) wireless network energy saving, and (6) the parameter settings calculator. Each of the six is discussed in more detail in this section.

MPEG-4 encoder.

It provides MPEG-4 code that is resistant to network fluctuations. It can easily be adapted to changing network conditions, such as bandwidth and packet drop probability. To counteract packet frames a gradual redefinition of the picture leads to a complete picture redefinition within n frames. Which means that after a frame drop, the video is restored within n frames. The impact of the various encoder parameters on the quality and on the resource usage has been thoroughly studied and corresponding models are available. (D2b, D2c, D2e, D7)

It answers to criteria 1, 2, 10, 13 –16. The BETSY function is the encoder. The stream parameters are Frame Size, Frame Rate, Intra-Period and Quantisation Parameter. The results are PSNR, and bit-rate. Given the display, the quality (framesize) can be adapted. Codec settings are related to power consumption. Given the packet losses, codec settings to optimize the quality are found.

BETSY Framework

It provides the software infrastructure to actually apply the BETSY results to a working prototype. A set of modules is defined which can be distributed over the network. A controller module contains the model and calculator facilities to recalculate the network and video codec as function of the network conditions, video equipment conditions and user wishes. The BETSY framework distributes the calculation results to the implicated devices. (D4b, D4c, D4d)

It answers to criteria 5 - 9, and 12. The demo allows to change codec settings and associated video quality, as function of changing bandwidth and power constraints. The Dependent on actual resource constraints and new resource constraints, the pareto algebra calculates a new setting, which is applied.

Power consumption models

They are the basis to calculate the influence of Frame Rate (FR), Frame Size (FS), Quantization Parameter (QP), and Intra Period (IP), on the power consumption of a given device. The impact of the same parameters as well as the packet size has also been modeled for the communication part. (D3b, D3c, D7)

It answers to criteria 12-16, and 18. It is possible to calculate power consumption for Access Point settings. For several devices, composed of more than one function, tables have been prepared to determine power consumption as function of Frame size, Frame rate and Quantisation parameter. Battery lifetime is changed by changing FS and FR such that power consumption is modified and battery lifetime adapted. Although the DOW promised this at function level, BETSY has provided these results for functions aggregated to the device level.

Video streaming delay calculations

We discern delays inside a device and delays over the network. Delays inside a device are calculated as function of the software structure inside.

Two device examples have been worked out: a MPEG-4 encoder and a software chain for a TV. The network delay for the energy saving wireless network on/off switching has been done. The delay over a chain is the composition of these individual delays. The influence of the appropriate MPEG-4 semantics, coupled to packet size, on the network delay is studied as well. (D2a, D2d, D2e, D7)

It answers to criterions 3, and 12. For a few specific but important cases within the BETSY consortium, delays over network and in device can be calculated quite extensively, but the two are naively composed.

Wireless network energy saving

The wireless network energy saving is based on a periodic switching on and off of the network interface cards orchestrated by the beacon signals, to be complemented with the channel aware rate control. (D2a, D2b, D2c, D2d)

It answers to criterions 6, 7, 12, and 13 (and indirectly 18). The settings of the card determine the bit rate, which has an effect on the quality of the video. For a specified quality it can be decided whether a network setting is sufficient.

Parameter setting calculations

A Pareto calculation approach has been used to specify the device settings dependent on network conditions and user wishes, such that energy conservative solutions are applied. (D7)

It answers to criterions 2, 3, 6, 9, 15-18. The pareto algebra is at the core of all the calculation to determine resource constraints and parameter values

Not addressed criterions

Not included are criterions 4, 11, 19. Perceived quality is out of the range of the BETSY project, because it needs quite sophisticated measurements and techniques involving groups of testing people. Fast bandwidth fluctuations can be handled to reduce their impact on the perceived quality, but nothing was done to estimate accompanying resource consumptions. There has been no effort to determine confidence intervals of energy consumption estimates.

5 Assessment of the BETSY objectives

This section lists the final objectives and their success criteria for the BETSY tasks. All information is stored in tables for readability and to assure a uniform presentation.

These tables contain the result of the self-assessment together with the evaluation criteria described in section 3.

5.1 Task 2.1 Identification of end-to-end delivery chain parameters

BETSY Task	Task 2.1 Identification of end-to-end delivery chain parameters
Task objective(s)	<ul style="list-style-type: none"> • Identification of components that contribute to end-to-end delivery chain. • Identification of attributes of these components, which describe the end-to-end delivery chain fully such that all relevant timing aspects can be quantified. • Relation between resource model and identified components is expressed
Success criteria	<ul style="list-style-type: none"> • All end-to-end delivery chain timing aspects can be described with the components and their attributes. • There is an unambiguous relation between the identified components and the resource models of wp3.
Yes: Complete success.	
<p>The timing aspects are not directly related to the components but to devices and network segments. In D7 the delay calculation for a CE device and the energy saving network solution is shown.</p> <p>The devices only contain the resource components identified by all wps.</p>	

5.2 Task 2.2 Stream description methods

BETSY Task	Task 2.2 Stream description methods
Task objective(s)	<ul style="list-style-type: none"> • Identification of relevant stream characteristics to put constraints on function parameters • Identification of relevant stream characteristics to efficiently allocate resources • Identify relevant semantic dependencies between video frames
Success criteria	<ul style="list-style-type: none"> • All stream characteristics that influence resource allocation and the parameter settings of the functions have been identified.
Yes: Complete success	
<p>The Frame Size (FS), Frame Rate (FR), the Quantization Parameter (QP), and the Intra Period (IP) are shown to be the most significant stream parameters that influence both the quality and the resource consumption, as established for different MPEG-4 encoder instances (in-house optimized MPEG-4 SP encoder, ASIC RTL simulations, VLC framework MPEG-4 encoder, TriMedia processor codec and Axis 207W wireless camera MPEG-4 encoder) in D2c, D2b and D7. Relations between stream characteristics and resource consumptions have been identified. The relations have been quantified with and without losses in the cited deliverables.</p>	

5.3 Task 2.3 Decomposition of end-to-end constraints

BETSY Task	Task 2.3 Decomposition of end-to-end constraints
Task objective(s)	<ul style="list-style-type: none"> • Construction of complete end-to-end chains composed of analyzable components • Identification of composable attributes of components such that attribute value of all components yields attribute values for whole chain • Method to calculate end-to-end timing characteristics over whole chain.
Success criteria	<ul style="list-style-type: none"> • End-to-end constraint calculation can be decomposed into tractable constraint calculations on components. • From calculation of constraints of components the end-to end constraints can be calculated.
No: partial success and fallback solution(s) implemented	
<p>The end-to-end constraints can be calculated with a Pareto approach, as shown in D7. However, the breeze components, identified as basic components, were not always amenable to analysis on existing devices. An aggregate of components on a given device was taken in that case. A good analysis of the components - identified for the Breeze- necessitates design approaches, which allow the analysis. It is questionable whether an energy aware design, that is amenable to analysis, will be recognized as sufficiently useful by the device manufacturers. We nevertheless do expect that designs, which allow energy optimization, will become more important in mobile application contexts. These designs will allow powering down of elements which are not necessarily the BETSY functions, but which can be used complementary to the BETSY results.</p> <p>The attribute values over the components can be composed. Delay is calculated in a convoluted way within a device and over a wireless channel. But individual delays can be composed rather trivially. Power consumption attribute is measurable over a device or BETSY functions. The attribute values and constraints can be composed with the pareto approach.</p>	

5.4 Task 3.1 Parameter and resource requirements

BETSY Task	Task 3.1 Parameter and resource requirements
Task objective(s)	<ul style="list-style-type: none"> definition of equivalence classes of resources defined in WP1 mapping of abstract resources to concrete resources
Success criteria	<ul style="list-style-type: none"> mapping and definition is complete and validated
Yes , complete success	
<p>The choice of BETSY functions in WP1 is instrumental to the success of the mapping. The functions were identified by selecting the functions, which were amenable to control by a user. Central to the choice was the single control point (e.g. remote control concept). For a representative set of devices the mapping from BETSY function to device has been done in D3a. The mapping is complete in the sense that the Devices selected for the mapping are representative for CE devices.</p>	

5.5 Task 3.2 Device functional component modeling

BETSY Task	Task 3.2 Device functional component modeling
Task objective(s)	<ul style="list-style-type: none"> Determine abstract resource requirements of components in chain Mapping of abstract resource requirements to physical resources
Success criteria	<ul style="list-style-type: none"> All abstract and concrete resources are identified and validated A complete and validated mapping exists
No : open ended problem	
<p>Abstract and concrete resources and a validated mapping exist as shown in D3b and D3c. It has not been shown that the mapping is complete or that all resources are identified for all possible designs of the future. However, we are complete with respect to the scenarios and the commercial devices, which play a role in the scenarios.</p>	

5.6 Task 3.3 Network functional component modeling

BETSY Task	Task 3.3 Network functional component modeling
Task objective(s)	<ul style="list-style-type: none"> • identification of devices in the network • mapping of physical devices to physical resources
Success criteria	<ul style="list-style-type: none"> • Representative devices are identified • A complete and validated mapping exists
No: open ended problem	
As for task 3.3, one example has been worked out completely in D3c. Looking in all possible network configurations was not attempted. The example is important enough to represent most network configurations in the home.	

5.7 Task 3.4 Model composition

BETSY Task	Task 3.4 Model composition
Task objective(s)	<ul style="list-style-type: none"> • Compose the resource requirements of the components, as done in task 3.3 and 3.2, to the resource requirements of complete devices and of a complete end-to-en video streaming chain.
Success criteria	<ul style="list-style-type: none"> • A complete and validated set of composition rules
Yes: Complete success	
Composition is shown in D7. It is complete in the sense that all components and resources, as identified in tasks 3.2 and 3.3, are composed. As a matter of fact, it is not possible to find a model for a component in an abstract way independent of the platform, but we are complete in the context of task 3.2 and 3.3 results.	
Actually we are more successful with respect to task 3.4 than with respect to task 3.2. We can provide concrete resource consumption for devices. As explained in D7, it is hard to measure and validate resource consumption functions for individual BETSY functions.	

5.8 Task 4.1 Specification of the BETSY Framework

BETSY Task	Task 4.1 Specification of the BETSY Framework
Task objective(s)	<ul style="list-style-type: none"> • Architecture of WP1 and model of WP2 are combined to a software framework • Interfaces of the components in the framework are specified • Interfaces are refined to support the timeliness and resource parameters of the components
Success criteria	<ul style="list-style-type: none"> • The required optimized settings defined in wp2 and wp3 can be realized with software framework. • Timeliness settings can be set and evaluated.
Yes: Complete success	
<p>Complete success with respect to the results of wp2 and wp3, as described in D4d. Timeliness depends strictly on the device design and network protocol design and is mostly fixed at design time. Once the first frame is displayed, the display times of all following frames are fixed as well. For example, we did not look at scenarios where videos are sent in advance to recuperate transmission delays for later frames. Delays are as small as possible with respect to the buffer space that is available, and the zapping time that needs to be minimized. Changing delays during video operation leads to unwanted hiccups and is avoided.</p>	

5.9 Task 4.2 Design of the test bed and the demonstrator

BETSY Task	Task 4.2 Design of the test bed and the demonstrator
Task objective(s)	<ul style="list-style-type: none"> • A set of components is identified to be realized on selected hardware and software platforms • The components realizable on an existing platform and the components realized by simulation are defined • Definition of quantifiable metrics for both simulation and selected platform
Success criteria	<ul style="list-style-type: none"> • The design must allow for real implementations as well as simulations of components in the test bed. • Provision of interfaces that promote the integration of components in a system and access to real components from application. • At least one complete processing chain (sender, receiver) of one scenario can be implemented on the demonstrator. • The design of the test bed provides for the implementation of additional scenarios, by using simulated components (e.g. WLAN network simulator).
Yes: Complete success beyond expectation	
<p>Three abstract descriptions of the BETSY scenarios are implemented. All scenarios can be executed with running code working on a real Breeze. No simulations are necessary. Code is described in D4c and D4d.</p> <p>Although the associated demonstrations are also demonstrated in other contexts, the BETSY framework goes beyond an ad-hoc demonstration, and shows a rigid framework on top of UPnP to which all devices adhere. It is the same framework that is used throughout the demonstrators. The shown demonstrations are just instances of the demonstrator.</p>	

5.10 Task 4.3 Reference implementation of the BETSY framework on the test-bed

BETSY Task	Task 4.3 Reference implementation of the BETSY framework on the test-bed
Task objective(s)	<ul style="list-style-type: none"> • implementation of task 4.2 design on test-bed
Success criteria	<ul style="list-style-type: none"> • Possibility to measure and set timeliness aspects and resource aspects for a single sender and single receiver interconnected by a network connection. • Possibility to simulate one scenario.
Yes: Complete success	
<p>Three scenarios are implemented. Timeliness remark is identical to task 4.2 evaluation. A measurement is done based on internal measurement of device during demonstration. All other measurements are done under controlled laboratory conditions and are difficult to reproduce during a demonstration.</p> <p>No simulation was needed. All scenarios were executed on the test bed demonstrator.</p>	

5.11 Task 4.4 Porting of reference implementation on the demonstrator set-up

BETSY Task	Task 4.4 Porting of reference implementation on the demonstrator set-up
Task objective(s)	<ul style="list-style-type: none"> • execution of task 4.2 and task 4.3 implementation
Success criteria	<ul style="list-style-type: none"> • Possibility to measure and set timeliness aspects and resource aspects for a single sender and single receiver interconnected by a network connection. • Possibility to simulate one scenario.
Yes: Complete success	
See remarks task 4.3	

5.12 Task 4.5 Validation and evaluation of BETSY models

BETSY Task	Task 4.5 Validation and evaluation of BETSY models
Task objective(s)	<ul style="list-style-type: none"> • Show that design can handle fluctuations in network conditions • Show that design can accommodate the sharing of resources by applications • Show that design can effectively trade battery consumption against quality
Success criteria	<ul style="list-style-type: none"> • All three objectives are met
Yes: Complete success	
<p>Design shows handling of fluctuation in network, explained in D4e. The demo involves walking away with a portable device, leading to a reduced bandwidth.</p> <p>Sharing of the resources can be accommodated by the design but is not demonstrated. Although not shown during the demos, it is possible to reduce the bit-rates of two sending videos such that they both fit within the available bandwidth. However, demonstrations are very much influenced by accompanying secondary effects, such as which bit rate still provides adequate viewing quality. How to accommodate fast fluctuations over the two senders. How to perceive reduced bandwidth and react adequately. How to calculate a fair bandwidth allocation, etc. etc. These questions are far beyond the scope of BETSY. On the other hand a receiving device (such as TV implemented with Trimedia in D3a), can share the CPU. By setting the CPU consumption via the BETSY framework, the trimedia device is shared and consequently the power is shared according to BETSY framework setting.</p> <p>Trading of battery consumption versus quality is shown. The battery level is switched down on sending device (helmet with camera), this is noted by the demonstrator software which adapts the Frame Rate and Frame Size to reduce the power consumption on sending device.</p>	

6 Conclusions

The set of criteria, defined for the acceptance of the technology developed during the BETSY project, is used for the evaluation of the results of the BETSY project.

In general we can conclude that the modeling of the BETSY components and devices has been successful in the context of the chosen scenarios and the associated devices. It has been more difficult than foreseen to relate the Breeze components to real measurements of devices to quantify the models. However, these difficulties clearly show where problems exist and what measures are needed to design more energy aware devices. Consequently, the models are less general than we originally aimed for. On the other hand, the applicability of the models to real cases was more successful than we anticipated.

The BETSY framework has been instrumental to the success of the project. It is possible to create a breeze and manipulate the parameters, which affect battery duration and video quality. No simulations are needed to show the breeze manipulation. However, the measurements themselves need to be executed in tightly controlled laboratory conditions. The results of the measurements are used inside the models, which we use to guide the parameter adaptations. Bringing all results together inside the BETSY framework necessitated to integrate the models. The end results show the feasibility of the BETSY approach.

References

- [1] Harmke de Groot et al., *Deliverable D1a, Scenario and use cases*, BETSY project, November 2004
- [2] BETSY project, *Description of Work*, May 2004