European Commission
Directorate-General Home Affairs
Prevention, Preparedness and Consequence Management of Terrorism and other Security-related Risks Programme

HOME/2009/CIPS/AG/C2-050

i-Code: Real-time Malicious Code Identification

D5.1: Midterm Management Report

Workpackage:	WP5: Project Management
Contractual delivery date:	June 2011
Actual delivery date:	July 2011
Leading partner:	FORTH
Contributing partners:	Politecnico di Milano, Institute Eurécom, Technical University of Vienna, Vrije Universiteit
Editor:	Evangelos Markatos
Contributors:	Sotiris Ioannidis, Stefano Zanero, Davide Balzarotti, Paolo Milani, Herbert Bos

Executive Summary:

In this deliverable we present the management report of the first year of the i-Code project. Overall, the project progresses well. All deliverables for this period have been delivered and the work on the remaining ones is well underway. The requirements have been collected, the proposed system has been designed, and the implementation has started. The partners have also published several papers in prestigious conferences and journals as can be seen at http://www.icode-project.eu/publications/

With the support of the Prevention, Preparedness and Consequence Management of Terrorism and other Security-related Risks Programme.
European Commission - Directorate-General Home Affairs

This project has been funded with the support of the Prevention, Preparedness and Consequence Management of Terrorism and other Security-related Risks Programme of the European Commission - Directorate-General Home Affairs. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Table of Contents

1 Publishable Summary

1.1 Summary of project objectives

The objectives of this project are: (i) to design and prototype a system for network-level real-time detection of malicious code spread, (ii) to customize and provide a malware infrastructure which will aid users to categorize and identify captured malware, (iii) to facilitate the detection of malware in high-speed next-generation networks through the design and prototyping of novel execution architectures, and (iv) to maximize the impact of the project through aggressive and effective dissemination of the project's results.

1.2 Work performed and results achieved so far

During the first year of the project, the partners successfully completed the requirements analysis and the design of the system. Both relevant deliverables have been produced and can be found at the web site of the project.

With respect to WP0 (Requirements Analysis), we provided an overview of the state of the art. Specifically we discussed background work done in the area of **high-speed pattern recognition**, **signature generation**, and **malware analysis**. These three areas form the backbone of real-time malware detection. We used this bibliographical survey as a guideline for the direction we should take towards developing a real-time malicious code identification framework. To further guide our compilation of the set of requirements necessary, we conducted interviews with experts in the field of network security. We selected those experts from a large cross-section of industry and academia, including ISPs, NRENs, security companies, search providers, CERTs, research institutions, etc. Then, we concluded our WP0 work by laying out the axes along which we plan to carry out the work necessary to build a real-time malicious code identification framework.

In WP1, we worked out an integrated design for i-Code console. i-Code brings together a variety of techniques for real-time detection and analysis of cyber attacks. Since the nature of these techniques varies wildly, we designed a solution that unites a number of different inputs in a meaningful way. The high-level i-Code design combines host and network level attack detection tools and various analysis techniques. Alerts are consolidated in a single interface, known as the i-Code console, to facilitate the administrator's tasks. We specifically want to provide for attack detection techniques in the network and on the host by means of network emulation (executing the payload of network traffic on the fly, and verifying whether or not it contains attack code), and behaviour-based detection (looking at the normal behaviour of applications in order to detect deviations that are likely to be caused by malware). For analysis, the project will provide a clustering technique to classify suspicious (shell)code. The console design consists of a viewer and a library of presentation functions—graphs, bar and pie charts, tables, etc.

1.3 Expected final results

During its second year, the project will proceed with the implementation, integration, and evaluation of the system. We expect that the proposed project will have significant impact to a variety of stakeholders including researchers, security practitioners, and network forensic departments. Indeed, in the short term, the project will offer the mechanisms and algorithms which will enable researchers and security practitioners to advance the state of the art in the area of network-level identification of malicious code. It will also offer

forensics departments a prototype toolset which could be used to pinpoint the effectiveness of existing tools and advance the accuracy of the real-time malware identification. In the medium to long term, the project will lay the missing foundations for the development of an integrated infrastructure for real-time detection of malicious code. The short and medium term impact of the project can be measured in several ways including the number of people who use the tools set, the number of people who have extended it, the number of publications which have cited it, etc.

1.4 Project web site

The web site of the project featuring all the public and dissemination information can be reached at http://www.icode-project.eu

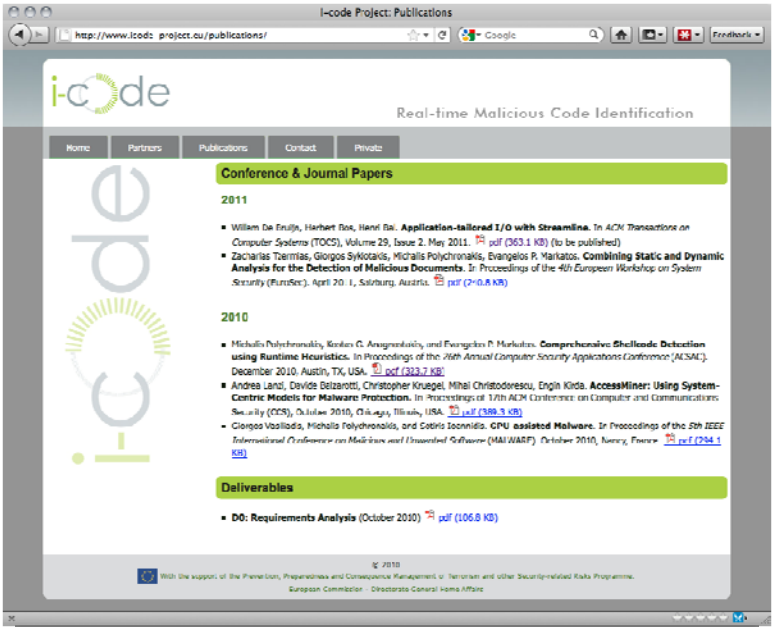


Figure 1: Screenshot of the website of the project. In this screenshot we see the papers published by the partners within the scope of the project.

2 Project Objectives, work Progress and Achievements, Project management

2.1 Project Objectives for the period

The main objectives for the reporting period are:

- To successfully complete the requirements analysis work (WP0)
- To successfully complete the design of the proposed system (WP1)

2.2 Work progress and achievements during the period

2.2.1 WP0: Requirements analysis

The goal of WP0 was to provide a requirement analysis with respect to real-time malicious code identification. The approach taken, followed a series of steps. We started by doing a state of the art bibliographical survey on the area of Internet attacks carried out by viruses, worms, and malware in general. This gave us a baseline from which to start from. Then, we contacted experts in the field and interviewed them. We selected experts from a large cross-section of industry and academia, specifically from ISPs, NRENs, security companies, search providers, research institutions, etc. This gave us the maximum possible coverage of expertise. We then processed to categorize the input we got and structure the requirements necessary for real-time malicious code identification. Using those requirements we made an initial proposal for the design that is going to be followed by the project.

2.2.1.1 Summary of progress towards objectives

The requirement analysis was conducted and delivered on time. A state of the art survey was conducted, experts were interviewed, the requirements were specified, and an initial design was proposed.

2.2.1.2 Highlight clearly significant results

We converged on six requirements for real-time malicious code identification, namely:

- **Speed**: we should try keep up with increasing network speeds
- **Coverage**: we must attempt to cover as many attacks as possible
- **False Positives**: care should be taken to address a potential high number of false positives
- **Device Heterogeneity**: we should take into account for new and upcoming devices
- **Network Heterogeneity**: we should consider different types of networks

- **Protocol Diversity**: we should consider the plethora of protocols being developed and deployed

Based on the above we broke up the preliminary design into four sections:

- Network-level emulation and communication pattern detectors
- Malware extraction and analysis
- High-speed detection
- Integration

2.2.1.3 Deviations from the plan and their impact

There were no deviations from the plan.

2.2.1.4 Reasons for failing to achieve critical objectives, if applicable

N/A

2.2.1.5 Use of resources

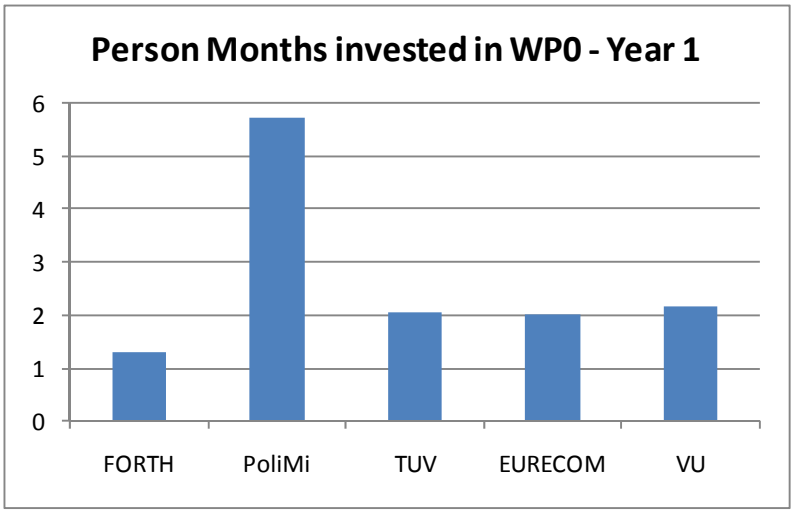


Figure 2: Person Months invested in WP0 during the first year of the project.

Figure 2 shows the number of person months invested in WP0 (Requirements analysis) during the first year of the project. We see that all partners have contributed to the work done.

2.2.1.6 Corrective actions

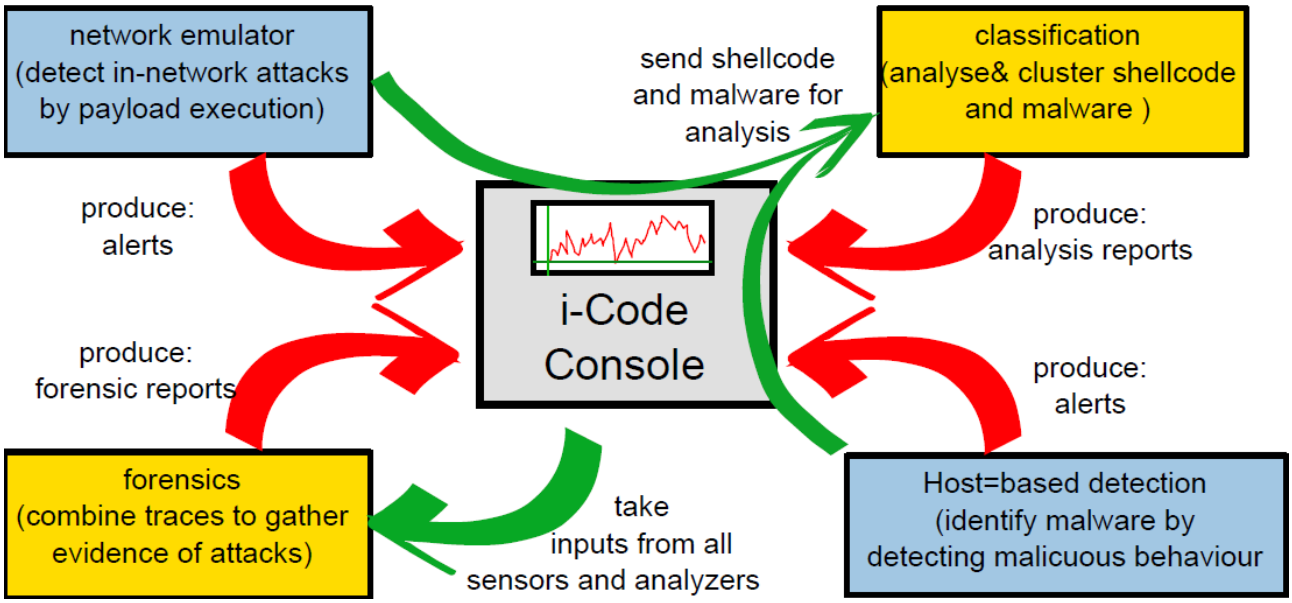
No corrective actions were required. This WorkPackage has been completed.

2.2.2 WP1: Design

In WP1, we worked out an integrated design for i-Code console. i-Code brings together a variety of techniques for real-time detection and analysis of cyber attacks. Since the nature of these techniques varies wildly, we designed a solution that unites a number of different inputs in a meaningful way.

1.1.1.1 Summary of progress towards objectives

Within the consortium, we have reached consensus about an overall design. The high-level i-Code design combines host and network level attack detection tools and various analysis techniques. Alerts are consolidated in a single interface, known as the i-Code console, to facilitate the administrator's tasks. We briefly summarise the design in terms of detection methods, analysis techniques, and presentation. The relations between these components are shown in the figure below.



Although we plan to make the console extensible so that others can plug in new tools, in our design we specifically want to provide for attack detection techniques in the network and on the host:

- Network emulation is a novel technique to detect an intrusion by means of executing the payload of network traffic on the fly, and verifying whether or not it contains code that looks like an attack (shellcode).
- Behaviour-based detection means that we look at the normal behaviour of applications in order to detect deviations that are likely to be caused by malware.

For analysis, the project will provide a clustering technique to classify suspicious (shell)code. By clustering shellcode, we can easily check whether something we detected is entirely new, or resembles code that we have seen before. As security software vendors receive many throusands of new samples each day, being able to separate the new ones from the known ones is increasingly important. The process of selecting what alerts to focus on, is known as triage. The i-Code console will help separate “serious cases” from “old news”.

The console design consists of a viewer and a library of presentation functions—graphs, bar and pie charts, tables, etc. By means of selection (e.g., ticking boxes), administrators using the console can select which data to represent and how to represent it. Thus we achieve the requirements of correlation and flexibility in presentation.

1.1.1.2 Highlight clearly significant results

We delivered D1, the system design.

1.1.1.3 Deviations from the plan and their impact

None.

1.1.1.4 Reasons for failing to achieve critical objectives, if applicable

N/A

1.1.1.5 Use of resources

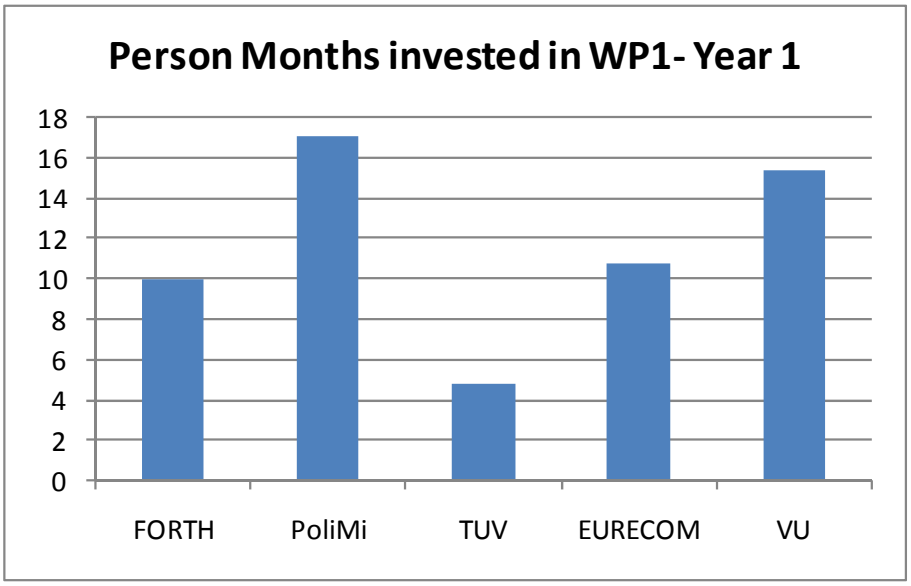


Figure 3: Resources invested in WP1 during the first year of the project.

1.1.1.6 Corrective actions

N/A

2.2.3 WP2: Implementation

This WP has not started during the reporting period.

2.2.4 WP3: Integration and Pilot Operation

This WP has not started during the reporting period.

2.2.5 WP4: Dissemination

In WP4 we have worked towards the dissemination of the project existence and of its result through various media (most significantly, through scientific, peer reviewed, publications).

2.2.4.1 Summary of progress towards objectives

During the period covered by this report, the i-Code consortium produced a total of 6 papers, all appearing in international conferences with peer-review.

The project also activated its website, collecting all of the public deliverables and the project publications. As discussed later, people from more than 60 countries have accessed the web site. Most of the accesses are from Europe and the States. This clearly shows that we are progressing against the objectives we set for WP4.

Our papers were discussed by the media (e.g. on the popular online blog "Slashdot" and on The Inquirer).

2.2.4.2 Highlight clearly significant results

The fact that some of our scientific contributions were picked up by mainstream media (such as "Slashdot" and The Inquirer) is a significant result.

D4.1 (the midterm report on dissemination activities) will be delivered alongside with this report.

2.2.4.3 Deviations from the plan and their impact

None.

2.2.4.4 Reasons for failing to achieve critical objectives, if applicable

N/A.

2.2.4.5 Use of resources

The following figure provides the number of person months invested in the project per partner. We see that the WP leader (PoliMi), the project coordinator (FORTH), and VU have invested the larger number of person months as expected.

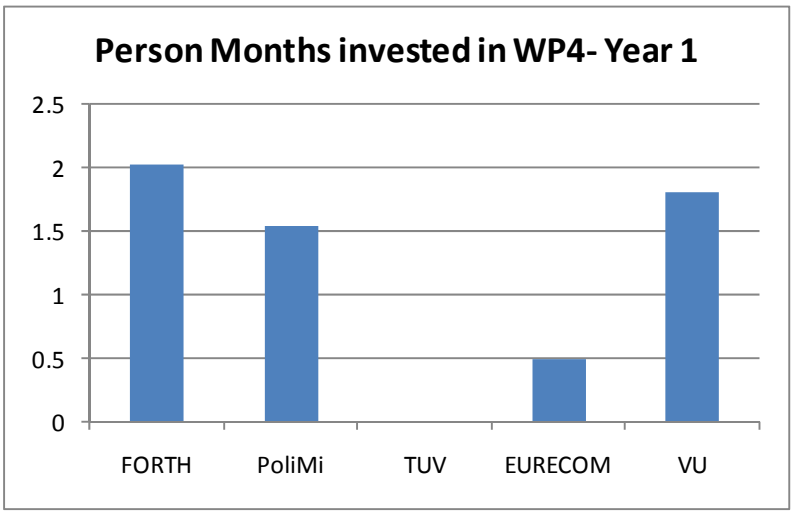


Figure 4: Person months invested in WP4 during the first year of the project.

2.2.4.6 Corrective actions

N/A.

2.2.6 WP5: Management

This WorkPackage started at the beginning of the project and will last for the entire duration of it. To avoid duplication of text, details on the Project Management WorkPackage (WP5) are given in the next section (section 2.3).

2.3 Project Management during the period

2.3.1 Consortium Management tasks and achievements

During the reporting period we successfully completed several management tasks including:

- **Consortium agreement**: We drafted and signed a consortium agreement which deals with various issues of the project, including IPR.
- **Meetings**: We held periodic project plenary meetings which were attended by all partners. The meetings were organized around an agenda circulated well in advance to all partners. During these meetings we discussed the progress of the tasks and scheduled the future work. After the meetings, the coordinator circulated the minutes containing the action points to all partners. During the reporting period we had three plenary meetings and one General Assembly meeting.
- **Collaborative Environment**: we operate on a 24/7 basis a collaborative repository based on SVN. Using this repository, partners can share documents and ideas. We also operate a mailing list for the project and individual mailing lists for the committees.
- **Committees**. We manned and started the operation of all project committees and bodies as mentioned in the proposal and subsequent contract. The meetings and attendance lists for these meetings can be found in the project's SVN
- **Reporting**. Prepared reporting templates for the partners to document their work, their person months and their expenses. The templates have to be filled twice per year.
- **Liaison**: The coordinator acted as a liaison between the partners and the commission conveying several questions as well as their replies.

2.3.2 Problems which have occurred and how they were solved

During the reporting period we did not encounter any problems.

2.3.3 Changes in the Consortium – if any

There were no changes in the consortium.

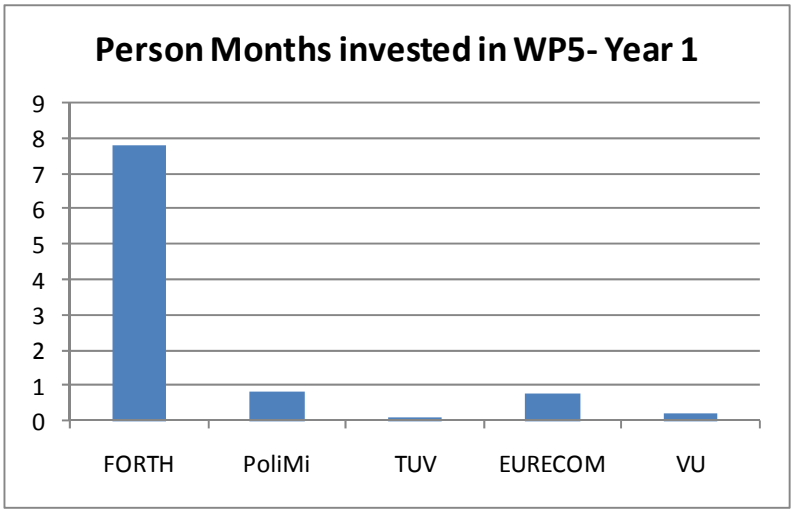


Figure 5: Person Months invested in WP5 during the first year of the project.

Figure 5 shows the person months invested in WP5 (Project Management) during the first year of the project. We see that the project coordinator (FORTH) invested most of the person months, while the rest of the WP leaders invested a small amount of capacity as well.

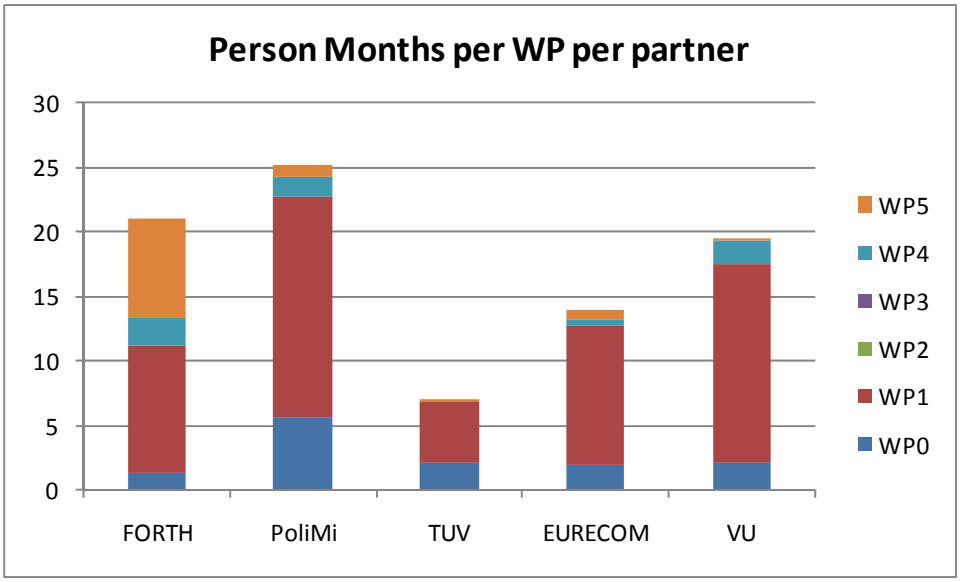


Figure 6: Person Months invested per WorkPackage per Partner during the first year of the project.

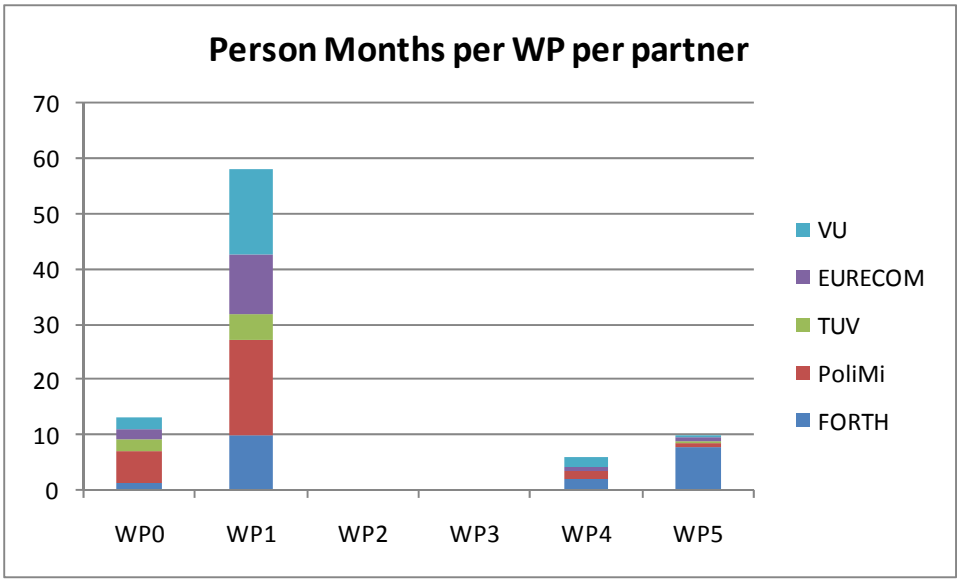


Figure 7: Person Months invested per WorkPackage per Partner. We see that the partners have invested their efforts in WP0 and WP1 which dominate the reporting period. WorkPackages WP2 nd WP3 will start in the next reporting period.

2.3.4 List of project meetings, dates and venues

During the reporting period the following project meetings were held:

1. First i-code plenary meeting, July 15th 2010, Amsterdam
2. Second i-code plenary meeting, December 2nd 2010, Milan
3. Third i-code plenary meeting, June 3rd 2011, Vienna
4. First i-code General Assembly meeting, Jun 2nd, 2011, Vienna

2.3.5 Project Planning and Status

The project proceeds as planned. During the first year, WorkPackages WP0 (requirements analysis) and WP1 (System Design) were completed. WorkPackages WP2 (Implementation), and WP3 (integration and pilot operation) are planned for the second year. WorkPackages WP4 (Dissemination) and WP5 (Management) started at the beginning of the project and are scheduled to run for its entire duration.

2.3.6 Impact of possible deviations from the planned milestones and deliverables, if any

We did not have any deviations from the planned milestones and deliverables during the reporting period.

In October 2011 Paolo Milani will be leaving TUV to the States. The partners discussed and proposed the following options to deal with the situation

- Option 1: Paolo to ask another person/Institute at TUV to take over WP2.
- Option 2: Transfer funds from TUV to VU or Eurecom to take leadership of WP2.
- Option 3: Invite another partner.

2.3.7 Any changes in the legal status of the beneficiaries

There were no changes in the legal status of the beneficiaries during the reporting period.

2.3.8 Development of the project website

The web site of the project has been operational since early August 2010. Figure 8 shows a screedump of the front page of the web site and Figure 9 shows the geographic distribution of the visitors of the pages of the web site. We see that people from more than 60 countries have accessed the web site. Most of the accesses are from Europe and the States.

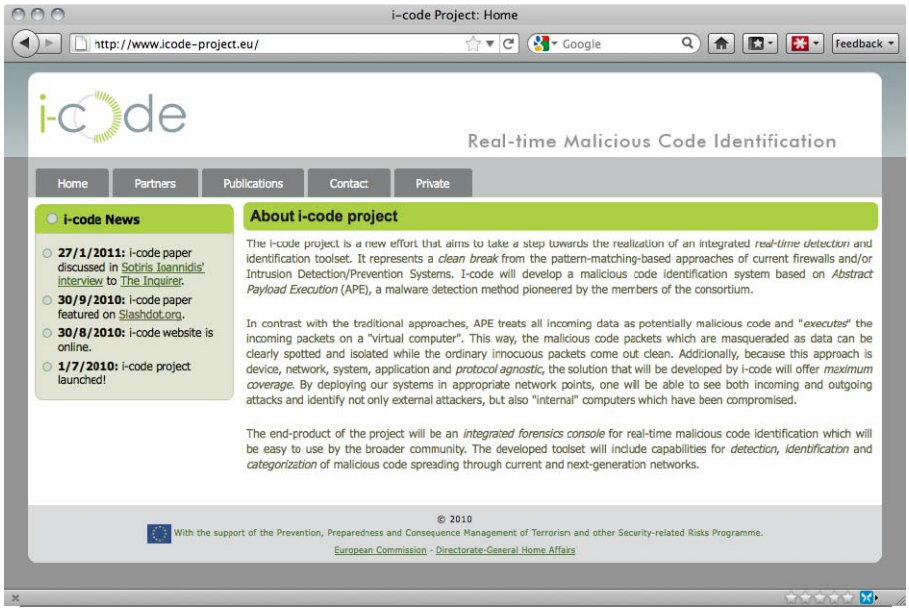


Figure 8: The web site of the project.

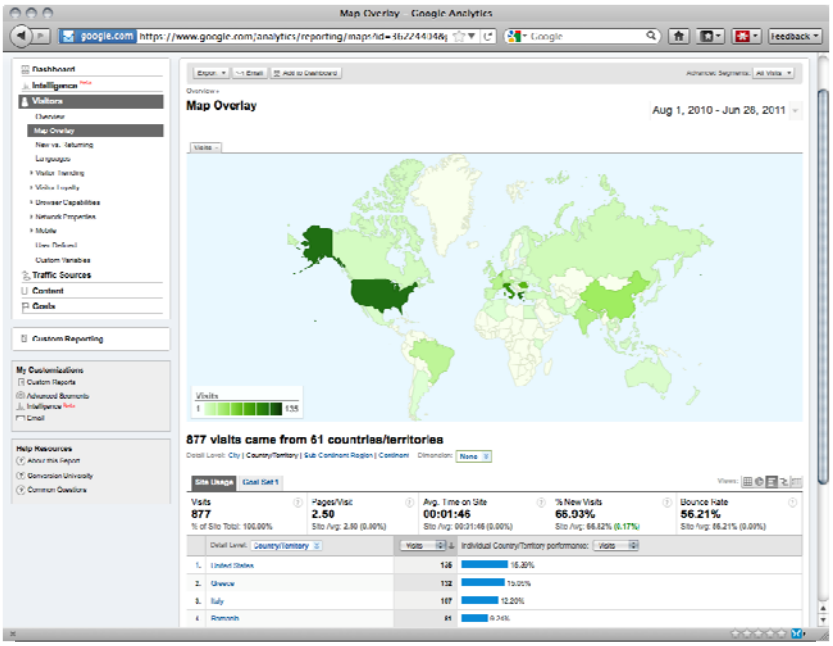


Figure 9: Visitors from more than 61 countries accessed the web site.

2.4 Deliverables and milestones tables

2.4.1 Deliverables

TABLE 1. DELIVERABLES

Deliverabe no.	Deliverable name	Version	Work Package no.	Lead beneficiary	Nature	Dissemination level[1]	Delivery date from Annex I (project month)	Actual/Forecast delivery date (project month)	Status No submitted/Submitted	Contractual Yes/No	Comments
D0	Requirements analysis	-	WP0	FORTH	R	PU	M3	M3	On the web	Yes	-
D1	System Design	-	WP1	VU	R	PU	M12	M13	On the web	Yes	-
D4.1	Midterm dissemination Report	-	WP4	PoliMi	R	PU	M12	M13	On the web	Yes	-
D5.1	Midterm Management Report	-	WP5	FORTH	R	PU	M12	M13	On the web	Yes	-

[1] **PU** = Public

PP = Restricted to other programme participants (including the Commission Services).

RE = Restricted to a group specified by the consortium (including the Commission Services).

CO = Confidential, only for members of the consortium (including the Commission Services).

Make sure that you are using the correct following label when your project has classified deliverables.

EU restricted = Classified with the mention of the classification level restricted "EU Restricted"

EU confidential = Classified with the mention of the classification level confidential " EU Confidential "

EU secret = Classified with the mention of the classification level secret "EU Secret "