www.wombat-project.org

European Commission
Seventh Framework Programme
Theme ICT-1-1.4 (Secure, dependable and trusted infrastructures)

ICT-216026-WOMBAT

Worldwide Observatory of Malicious Behaviors and Attack Threats

Requirements Analysis and Specification

Workpackage:	WP2
Deliverable:	D05 (D2.3)
Date of delivery:	30/06/2008
Version:	Final
Responsible:	NASK
Authors:	NASK with contribution from: FORTH, TUV, VUA, POLIMI, FT, EURECOM, HISPASEC, SYMANTEC
Data included from:	FORTH, TUV, VUA, POLIMI, FT, EURECOM, HISPASEC, SYMANTEC
Contact:	Elzbieta.Nowicka@cert.pl Piotr.Kijewski@cert.pl

Executive Summary

This document outlines the requirements for early warning systems built on technology provided by the WOMBAT project, setting out both: functional and non-functional requirements. The collected requirements reflect the identified user needs and the key directions to be followed within the research and development Work-packages (WP3-Data Collection and Distribution, WP4-Data Enrichment and Characterization, WP5-Threat Intelligence).

The document starts from an assessment of user requirements gathered from potential users including external participants in the Closed Workshop and the WOMBAT development group. This part covers expectations of distinct classes of data users such as: security vendors, malware researchers, ISPs, CERT teams, Government, financial institutions and home users. It details the requirements for the system architecture, data and system functions, and specifies performance, availability and security features to provide sufficient functionality. It also defines user interface, testing and configuration management requirements.

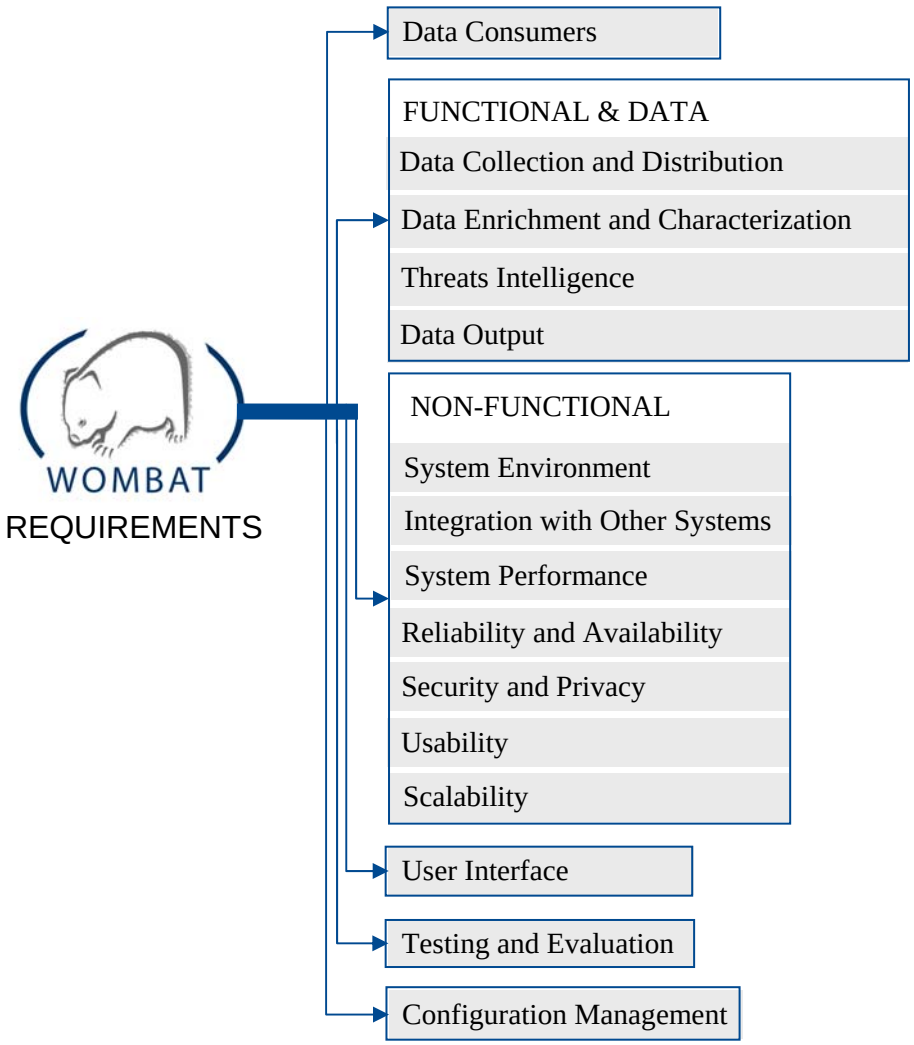
TABLE OF CONTENTS

1 INTRODUCTION

1.1 Scope

This is the requirements specification document for the WOMBAT system. Its purpose is to provide a collection of statements to form research directions for the WOMBAT project. The requirements specified here are based on several inputs: Description of Work (DoW) document [5], outcome from the Closed Workshop (April 21-22, Amsterdam), input from several informal discussions among the project consortium as well as opinions and expectations of potential WOMBAT users (i.e.: ISPs, CERTS, antivirus companies, security researchers, security-conscious organizations and home users). This document covers users, functional, data, non-functional as well as testing and configuration management requirements. At the same time, this document is intended to specify a kind of a "road map" that gives descriptive research directions for the project. Therefore, this document is not intended to supersede the DoW, but serve as a reminder of potential consumer expectations. It will help us to focus on providing solutions that will attempt to address in the best way the ideal functionality that is expected for such a system.

1.2 Requirements Taxonomy



1.3 Requirements Prioritization

For the purpose of requirements prioritization, within the entire document we distinguish the following classes of requirements:

- ESSENTIAL
- DESIRABLE
- OPTIONAL

This document specifies research solution requirements as a combination of prior agreed requirements specified in DoW document (as deliverables) and a collection of expectations of WOMBAT potential users, including comments received from the participants of the Closed Workshop in Amsterdam (April 21-22). We will interpret the above classes of requirements as follows. The "ESSENTIAL" requirements are critical points that must be accomplished by an operational system based on WOMBAT-like components. In this spirit, requirements specified according to DoW are obligatory. However, for the remaining “ESSENTIAL” requirements, given the time and cost constraints of such a research project, we envision that some of the components developed or the overall integration may not meet these requirements, particularly when stability and performance are concerned. Any requirement classified as "DESIRABLE" would enhance the system, but is not essential for the project. An "OPTIONAL" requirement is facultative for the project.

***Note*:** Requirements taken directly from DoW document are shadowed.

1.4 Document Overview

This section provides an overview of the entire document. This document describes data, functional and behavioral requirements for the system to be developed within the WOMBAT project. This document is structured as follows. *Chapter 1* gives brief information about the scope of this document. Also, it provides the taxonomy of requirements and the requirements prioritization method. *Chapter 2* provides characteristics of main end users of the new system. This chapter also lists existing systems that will provide main data input for the system. Additionally, it specifies assumptions and some constraints on the system development. *Chapter 3* characterizes targeted audience and their expectations from a new system formalized as user requirements. *Chapter 4* defines data requirements as well as operational and functional requirements for the activity of the system, including: the system architecture, requirements for design and deployment of its sensors, the type of information that has to be acquired from different kinds of sensors, and requirements for data repository, as well as requirements for the results of data enrichment and threat intelligence processes. *Chapter 5* specifies non-functional requirements, i.e. constraints on the system design and implementation. *Chapter 6* specifies requirements for API of the WOMBAT system. *Chapter 7* proposes requirements for system testing and evaluation. *Chapter 8* describes requirements for WOMBAT configuration management.

2 GENERAL INFORMATION

2.1 Users Characteristics

Main end users of the WOMBAT system will include:

- **Security vendors**: auditing and consulting services, provide and develop anti-malware and other computer security solutions and tools (products), malware and vulnerability analysis, malware collection; (representation by Symantec, Hispasec Sistemas)
- **ISPs**: provide consumers or businesses with access to the Internet and related services, provide web hosting, domain name registration, collocation, Internet transit, also provide security issues to their customers; (representation by France Telecom, NASK)
- **CERT teams**: respond to security incidents occurring in the Internet, cooperate with other CERTs and ISPs, provide secure contact to report an incident, analyze the state of the Internet security, provide incident reaction and prevention, provide security information and warnings as well as education and training; (representation by NASK/CERT Polska)
- **Banks**: provide financial services to their customers via Internet: banking, investment, brokering, etc.; (representation by Clearstream, EAB)
- **Government**: regulations about telecommunication and the Internet, provide public information and government office services to citizens via webpage;
- **Researchers**: research and teaching (education), development of ideas related to a broad range of theoretical and practical aspects of computer security and privacy issues (Internet threat analysis, intrusion detection, cryptography, etc.); (representation by France Telecom R&D, Institut Eurecom, NASK, FORTH, Politecnico di Milano, Technical University Vienna i Vrije Universiteit Amsterdam)
- **General public**: theoretical customers of most of above users (especially security vendors, ISPs, banks, government – suppliant), typical website viewers and Internet surfers.

2.2 Input Systems

Basic sources of information about threats and malicious events used in the WOMBAT project will include:

- **DeepSight**, Symantec

The Symantec DeepSight Threat Management System and Symantec Managed SecurityServices [4] consists of more than 40,000 sensors monitoring network activity in more than 180 countries and comprehensively tracks attack activity across the entire Internet. Additionally, Symantec gathers malicious code data along with spyware and adware reports from over 120 million client, server, and gateway systems that have deployed Symantec's antivirus products and opted into sharing such reports within the agreed upon terms of privacy and anonymization. Some information delivered by the DeepSight will be shared within WOMBAT project through an XML-based proprietary API.

- **Leurré.com (SGNET)**, Institut Eurécom

The Leurré.com project [8] operated by Institut Eurécom is based on a broad network of honeypots covering more than 30 countries. The architecture consists of a distributed

network of low-interaction honeypots (based on honeyd), medium-interaction honeypots (based on the ScriptGen technology in order to enrich the network conversations with the attackers) and a central server. Honeypots of each partner monitor three unused IP addresses. All traces captured on each platform are uploaded on a daily basis into a centralized central relational database. Some data delivered by system will be provided to WOMBAT.

- **Argos**, VUA

Argos [3] is a full and secure system emulator designed for use in honeypots. It is based on Qemu, an open source emulator that uses dynamic translation to achieve a fairly good emulation speed. Argos extends Qemu to enable it to detect remote attempts to compromise the emulated guest operating system. Using dynamic taint analysis it tracks network data throughtout execution and detects any attempts to use them in an illegal way. When an attack is detected the memory footprint of the attack is logged. Activities in Argos sensors can be captured and analyzed in the context of WOMBAT.

- **Honey@Home**, FORTH

Honey@Home [6] is a honeypot-based system designed to gathering and analyzing information on cyber-attacks that uses home-users hosts as sensors. It is designed to manage and lightweight on system resource usage. Honey@Home forwards traffic to unused IP addresses or ports of home-user host to a honeypot farm and forwards the replies back to the attacker. It runs in the background of a home-user computer.

- **Anubis (Analyzing Unknown Binaries)**, International Secure Systems Lab (Vienna University of Technology, Eurecom France, UC Santa Barbara)

Anubis [1] is a tool for analyzing the behavior of Windows executables with special focus on the analysis of malware. To this end, the binary executable is run in an emulated environment and its (security-relevant) actions are monitored. The generated report includes detailed data about modifications made to the Windows registry or the file system about interactions with the Windows Service Manager or other processes and it logs all generated network traffic. In the context of WOMBAT, this tool will help to characterize malware and will be useful in the process of Threat Intelligence.

- **ARAKIS**, NASK / CERT Polska

ARAKIS [2] is the nation-wide early warning system built by CERT Polska that collects and correlates data from a wide variety of sources including low-interaction honeypots, firewalls, antivirus systems and darknets. The system is oriented towards detection and characterization of new threats based on the automated analysis of captured honeypot payloads and supporting data from other sources. Activity observed by ARAKIS's sensors and analysis performed by the system can be partially used by WOMBAT via API.

- **VirusTotal**, Hispasec

VirusTotal project [9] offers a free service for scanning suspicious files using several antivirus engines. Companies, institutions, organizations and individuals can submit malware samples that are scanned by the VirusTotal service using more than 30 different antivirus products. The number of malware samples received by the service is more than twenty thousands per day. This malware collection will be very useful for WOMBAT.

2.3 Assumptions, Dependencies and Constraints

There are a number of critical issues which ESSENTIAL be either considered during the early stages of the project development or addressed at a later phase to make reaching the goals of the project feasible. They include:

- **Data collection, access to data, requests for data**

Data collection is the starting point for further analysis. One potential major obstacle may be the lack of completeness of the collected data. This may be the result of (a) legal issues that force original sources to anonymize parts of the data offered, (b) weaknesses of the original data collection systems or (c) problems related to coverage. One concern is where to put the collection points for a better coverage. Also, collection points may be detected by attackers and avoided by them. Moreover, using sensors that operate only on honeypots (regardless of whether they are server-side or client-side) may result in a failure of detection and collection of attacks, as honeypot configuration is usually different from that of production systems. User behavior may also play a critical factor in enabling an attack to become feasible, something not possible to reproduce on a honeypot system.

Although aggregated and anonymized data would be valuable and easier to share, there is a strong need for raw data to enable meaningful research, especially for the purpose of scientific and multi-perspective analysis.

Sharing data is the core problem. For the global collaboration within the project, there is a need to establish both formal and semi-formal agreements between institutions (or individuals) that will cover privacy and legal issues concerning exchanging data and using it. In this case, any limitations on the use of the data concerning privacy issues should be identified and documented.

- **System architecture**

The basic structure and type of WOMBAT architecture which is proposed in section 4.1 will evolve, with changes depending on the data shared in the system.

In the data collection infrastructure some data will be held centrally (under a central agreement), while other data will be held by individual organizations under separate agreements.

Centralization or standardization of formats would be difficult to introduce. A common API requires consensus between safety and flexibility. For the system to scale in the future, standardization is necessary. However, a ‘start small, think big’ approach is more practical.

Scalability is in fact a critical issue at all levels of the WOMBAT system, as the wealth of collected information will be huge, making analysis time-consuming and heavy resource wise.

- **Technology and tools**

To obtain meaningful research results, there is a need for adequate analysis techniques. The lack of advanced technology and appropriate analysis tools would limit the scope of research and ability to obtain meaningful results. Code analysis is a hard problem. Also, the true heads behind malicious activity may hide behind layers of indirection that makes it difficult to get to them.

A major challenge is to provide a translation mechanism to interpret the collected data from WP3 to the model of WP4. It will be important to address the completeness and soundness of the translation. Other questions must also be taken into account, including: whether the translation can be totally automatable or requires some configuration.

One of the major assumptions for the WOMBAT system is that most possible behavior patterns can be reliably extracted from malware, so that subsequent algorithms get to operate on meaningful data. Problems here may include anti-analysis methods employed in malware, such as the detection of an emulator. Resilience to such methods should be taken into account at an

early stage of system development. At the data collection level, specific system configuration or user interaction may make malware function in a different manner, making the extraction of such behaviors difficult in an artificial environment.

- **Testing**

Testing methods and tools should be created for independent assessment of security software. Most tests and in particular for AV products are not really independent making the result less trustworthy and quite variable. Moreover, actual test procedures simply confront the products to a huge pool of known malware and deliver percentages thereby solely assessing signature-based detection. More complete assessment should consider additional aspects such as resilience to unknown malware which is also true for the results of the project [7].

3 DATA CONSUMERS REQUIREMENTS

This chapter characterizes targeted audience and their expectations for the WOMBAT system. Among the project's targeted audience there are several distinct groups (such as security vendors, malware researchers, ISPs, CERTs, Banks, Governments and others) with different expectations for the system. The following requirements, which are specified separately for each distinct group of data users, are based on the presumed usage of the new system. They are also the basis for the specification of the most of the functional (Chapter 4) and partially also of the non-functional requirements (Chapter 5). Among functional requirements, Section 4.1 defines various kinds of data collected from different sources – these correspond to the data requirements explicitly listed in this chapter and/or provide basis for the further analysis. Section 4.2 specifies the new system functionality which satisfies many of the requirements listed below that are not satisfied by raw data and basic statistics. Section 4.3 refers to the most advanced requirements defined below. Requirements included in this chapter, which do not request specific KID (Knowledge, Information, Data), concern presentation and addresses data output (Section 4.4) as well as some of the non-functional requirements.

Requirements included in the following tables reflect presumed expectations from the WOMBAT system of its potential users. These requirements are the result of the WOMBAT consortium members' research, informal interviews with representatives of some of the user groups and input from the closed WOMBAT workshop.

3.1 Security Vendors and Malware Researchers

Table U1. Requirements for the WOMBAT system
from the point of view of security vendors and malware researchers

NO.	USER REQUIREMENTS DESCRIPTION	PRIORITY
[U1-1]	Providing access to malware samples (for selected users only) and enable sharing of such samples based on agreed upon access procedures.	ESSENTIAL
[U1-2]	For a given malware sample, providing any available metadata, including descriptions, geographical statistics, time, etc. as well as – when available – analysis logs with recorded system calls, their arguments and collection point (for malware analysts)	ESSENTIAL
[U1-3]	Allowing for automated signature generation methods	OPTIONAL
[U1-4]	Acting as part of the users threat collection infrastructure	OPTIONAL
[U1-5]	Allowing for the possibility of searching for information for a given malware sample based on basic characteristics (MD5/SHA, file length, simple behavioral characteristics like port numbers, etc.)	ESSENTIAL
[U1-6]	Allowing for threat intelligence analysis enabling identification of root causes of attacks and prediction of attack vector changes	DESIRABLE
[U1-7]	Providing feedback about any identified false positives to the original information source	OPTIONAL

3.2 Internet Service Providers

Table U2. Requirements for the WOMBAT system from the point of view of ISPs

[U2-1]	Providing information about current threats useful from the point of view of client support (for example customer call centers), particularly results of semi-automatic malware analysis including but not limited to infection symptoms, known malware removal procedures, information about patches blocking the vulnerabilities used by the threat and workarounds if patches are not available	ESSENTIAL
[U2-2]	Providing threat signatures to enable filtering of known malicious traffic	DESIRABLE
[U2-3]	Enabling users to place sensors within their own networks to observe statistics of attacks that threaten their own customers and to gain additional knowledge about those threats from the threat analysis and intelligence carried out by the system	OPTIONAL
[U2-4]	Providing port activity and other statistics, if collected (netflow records, …)	ESSENTIAL
[U2-5]	Provide pro-active protection measures and self-care cleaning support for ISP customers	DESIRABLE
[U2-6]	Measure and assess the impact of the threat on real-time traffic such as VoIP and IPTV	DESIRABLE
[U2-7]	Measure and assess the impact of the threat on the networking infrastructure (routers, switches, firewalls, DSLAMs, BRAS, SBCs, …) and associated services (RADIUS, DNS, DHCP, …)	DESIRABLE
[U2-8]	Measure and assess the impact of the threat on boxes (e.g. ADSL home routers) and terminals (e.g. phones)	DESIRABLE

3.3 CERTs

Table U3. Requirements for the WOMBAT system from the point of view of CERTs

[U3-1]	Providing information about attacks and malware originating from or targeting the IP range of the CERT constituency	ESSENTIAL
[U3-2]	Providing threat signatures accessible in an automated way (API)	OPTIONAL
[U3-3]	Providing threat statistics, including port activity	ESSENTIAL
[U3-4]	Providing information of any significant correlation identifying groups operating in the user's area	DESIRABLE
[U3-5]	Providing early warning about new identified threats, both malware and exploits	DESIRABLE
[U3-6]	Enabling tracking of activity of malicious groups using the threat intelligence capability and enhancing digital forensics results with correlated results from other analyses	OPTIONAL
[U3-7]	Support information exchange with peers (other CERTs, FIRST, …)	DESIRABLE

3.4 Banks

Table U4. Requirements for the WOMBAT system from the point of view of Financial Institutions

[U4-1]	Providing information about any malware specifically targeting the user or his clients, including results of semi-automated analysis. This information will be based on profiles that would have to be supplied. Profiles could be, for example, IP ranges or domain names.	ESSENTIAL
[U4-2]	Providing information about any known phishing attempts targeting the user's clients	DESIRABLE
[U4-3]	Providing information about any correlations between different malicious activities identifying groups engaging in phishing targeting banks and/or CC number theft, warning the user about any new identified activity of such groups	OPTIONAL
[U4-4]	Providing threat signatures for the user's network devices, most importantly including signatures of threats against HTTP servers used in e-banking	OPTIONAL
[U4-5]	Enabling checking for malware-infected or otherwise suspicious pages on the user's site, active alerting is preferred	DESIRABLE
[U4-6]	Enabling finding other banks targeted by the same group to allow joint action against the attack	OPTIONAL
[U4-7]	Providing information about new vectors of attack	DESIRABLE

3.5 Government

Table U5. Requirements for the WOMBAT system from the point of view of Government

[U5-1]	Providing information about malicious behavior targeting national security of any country (cyber-terrorism)	DESIRABLE
[U5-2]	Providing proper authentication of data sources	ESSENTIAL
[U5-3]	Providing available information about groups behind the malware, including probable locations (using information such as IP address statistics, languages, etc.)	DESIRABLE
[U5-4]	Enabling checking for any known phishing attempts or malware infections on governmental sites	DESIRABLE
[U5-5]	Providing information about malware and attack attempts against Government sites. Government institutions interested in getting notifications should provide the project with patterns to be detected.	DESIRABLE
[U5-6]	Allowing reports about general trust in e-commerce, general levels of phishing attempts, id theft etc	DESIRABLE
[U5-7]	Providing information about new vectors of attack	DESIRABLE

3.6 Business Users (Network and Systems Managers) / Administrators

Table U6. Requirements for the WOMBAT system
from the point of view of Business Users/Administrators

[U6-1]	Providing information about malware and any available metadata for malware samples, including descriptions, malware-URLs (for example, for proxy configuration) geographical statistics, time, etc.	ESSENTIAL
[U6-2]	Enabling checking for malware-infected pages on the user's site	DESIRABLE
[U6-3]	Enabling searching for known attacks originating from the user's IP range	DESIRABLE
[U6-4]	Enable searching for known attack targets	DESIRABLE

3.7 General Public

Table U7. Requirements for the WOMBAT system
from the point of view of General Public

[U7-1]	Providing basic threat statistics in a visually entertaining and educating way to increase public awareness of network security issues	DESIRABLE
[U7-2]	Providing TOP-X lists of most exploited vulnerabilities	DESIRABLE
[U7-3]	Providing threat signatures for home network devices, including WiFi, ADSL, IPTV and VoIP	OPTIONAL
[U7-4]	Providing available threat signatures for software such as ClamAV	OPTIONAL
[U7-5]	Provide and/or support self-care malware removal tools	OPTIONAL

4 FUNCTIONAL AND DATA REQUIREMENTS

In this chapter, we define data requirements as well as operational and functional requirements for the activity of the WOMBAT system. First three sections of this chapter correspond directly to Workpackages WP3, WP4 and WP5 of the WOMBAT's DoW document. The last section of this chapter refers to the output of the WOMBAT system, i.e. information that is expected to be generated as the result of the carried research.

4.1 Data Collection and Distribution

This section refers to the WP3 Work-package and describes: the type of the WOMBAT architecture, requirements for its sensors design and development, kind of data to be acquired from sensors. The objective is to improve malware samples collection.

4.1.1 Architecture of the Infrastructure

The WOMBAT infrastructure will be based on the various existing sensors developed within previous projects, including honeypots (Leurre.com, VirusTotal, NoAH) and attack detection systems (DeepSight Threat Management System, ARAKIS) (see Section 2.3, basic resources) as well as new sensors implemented and deployed within the WOMBAT project (see Section 2.3, external resources). Existing sensors are, however, mostly passive ones, while new sensors will include mid-interaction honeypots, such as web crawlers that actively seek malware on the Internet, and Scriptgen/Argos solutions. New sensors will also include wireless and Bluetooth sensors.

The basic WOMBAT infrastructure will consist of the following elements:

Input interface: to which sensors will feed their alert data and will in turn be available to authorized entities that request them. In order to preserve the architecture and functionality of the currently deployed sensors, these sensors will be offered the choice of selecting how they interface into the presentation architecture. The possible choices include:

- Email - data can be sent via email to an mail address dedicated for collecting alert notifications from monitoring sensors.
- FTP/SCP - data can be uploaded to a central repository.
- Web Service - data can be uploaded via a request to a web service developed for the purpose of gathering alert data.
- HTTP - data can be uploaded by simply using HTTP requests.

For any of the existing sensors that has any other (reasonable) way to distribute data, the client part of could be additionally developed.

Database management system: (e.g. mysql) to store in a centralized way the alert data gathering from the existing sensors. The database will hold all the data and metadata collected. Because of the heterogeneity of the data collected by the existing sensors extra caution must be taken during the design of the database schema in order to be easily extensible. In addition, the database schema will also take into consideration the fact that new forms of data may appear during the WOMBAT project.

Communication channel with the client applications: this outgoing channel, as opposed to the incoming ones, will have only one form (e.g. web service, HTTP). In addition to communication, this layer will also take care of the different user roles, as described in the draft, which are home users, security vendors, ISPs, etc.

GUI for the WOMBAT's database: the formal GUI for the database will be in the form of a web interface. It will provide the users with many visual representations of the current threats on the Internet.

4.1.2 Data Sensors Design and Deployment

The following requirements concern deployment of existing sensors as well as design and deployment of new sensors that will provide data to the WOMBAT system.

Table R1. Requirements concerning WOMBAT sensors

NO.		SYSTEM REQUIREMENT DESCRIPTION	PRIORITY
[R1-1]		There will be a specification of the interface between WOMBAT and other input (basic and external) systems (see 2.3).	DESIRABLE
[R1-2]		Communication layer between WOMBAT and existing individual sensors will be based on that sensors' native data distribution protocol (if practical and not overly complicated). For example, methods for transferring data will include: email, ftp, scp, http, etc.	DESIRABLE
[R1-3]		If several protocols are available for a given sensor and at least one of them is also used by other sensors in WOMBAT, then that protocol will be used to minimize implementation effort.	DESIRABLE
[R1-4]		New sensors will have a unified way of distributing data (e.g. web services).	DESIRABLE
[R1-5]		If one of the protocols used by existing sensors meets the needs of new sensors, then that protocol will become the standard.	DESIRABLE
[R1-6]		Some protocols are common and will be supported by WOMBAT:	
	a.	Email - data can be sent via email to an e-mail address dedicated for collecting alert notifications from monitoring sensors.	DESIRABLE
	b.	FTP/SCP - data can be uploaded to a central repository.	DESIRABLE
	c.	Web Service - data can be uploaded via a request to a web service developed for the purpose of gathering alert data.	DESIRABLE
	d.	HTTP - data can be uploaded by simply using HTTP requests.	DESIRABLE
[R1-7]		Sensors using other protocols will be accepted and the client part of the interface will be provided by WOMBAT, subject to practicality constraints.	OPTIONAL
[R1-8]		The reason why data was collected will be clearly identified by WOMBAT [AV detection | honeypot | suspicious-behavior | suspicious source | ...]	DESIRABLE
[R1-9]		WOMBAT will include sensors that work on production systems.	OPTIONAL
[R1-10]		WOMBAT will have an assessment that determines whether something developed within a project (e.g. that provides data to the repository) is sufficiently mature to be released publicly.	DESIRABLE
[R1-11]		**Improving tandem crawler technology** (i.e. improving the quality and capabilities of client-side threat sample collection technologies of honey-crawlers):	
	a.	Improving metrics for suspicion in behavioral deviations between infected and clean crawlers	ESSENTIAL

	b.	Improving metrics for confidence in benign deviations among clean crawlers	ESSENTIAL
	c.	Refining management framework to improve consistency of behavior among clean crawlers	ESSENTIAL
	d.	Leveraging machine learning techniques to more effectively prioritize potentially malicious deviations	ESSENTIAL
	e.	Leveraging machine learning to more reliably and more accurately cluster similar deviations	ESSENTIAL
	f.	Leveraging background and side-ground technologies and infrastructures to more effectively target the crawler toward frequently malicious sites or generally benign sites as needed	ESSENTIAL
	g.	Re-architecting the software for efficiency to compress (five) physical machines into virtual machines that fit within reasonable desktop computing hardware so that the technology is easily used by partners	ESSENTIAL

4.1.3 Input Data and Information

The WOMBAT system is intended to collect the wide diversity of data with as much details as possible to provide the meaningful and multi-perspective analysis of different Internet threats. Thus, among data and information to be gathered from WOMBAT sensors there will be: collections of malware, logs from firewalls and IDS sensors, honeypot-based information, darknets, alerts from early warning systems, and also (however considered as a future input) information from mobile devices and RFID. The following tables define requirements to characterize features to be provided upon particular data collections.

Table R2. Malware Collections

[R2-1]	Information acquired from malware collections	ESSENTIAL
[R2-2]	Malware file sent to WOMBAT should be packed & protected (example: in a ZIP file with password 'infected', an industry non-official standard)	ESSENTIAL
[R2-3]	Provide metadata: hashes of the original malware file [MD5 & SHA1|SHA256] (not only MD5). This information may be useful even without actual samples if enough information is present to identify the sample (hashes, filesize).	ESSENTIAL
[R2-4]	Provide timestamp of collection	DESIRABLE
[R2-5]	Provide source id should include some degree of detail (keeping it anonymized) about the source of that sample, even in general groups like 'trusted malware researcher', 'CERT', 'honeypot', 'user', etc	ESSENTIAL
[R2-6]	Provide basic metadata of the sample including original filename, if renamed	DESIRABLE
[R2-7]	Provide source / infection vector [www | e-mail | p2p | document | exploit | bluetooth|...]	DESIRABLE
[R2-8]	Provide reason for being collected [AV detection | honeypot | suspicious-behavior | suspicious source | ...]	DESIRABLE
[R2-9]	Provide the associated information, if any (example: AV detection [engine-name, version, malware-name], www [URL]))	DESIRABLE
[R2-10]	Contain extra info given by certain tools and AV products (i.e. Norman Sandbox reports, PE structure info, etc)	OPTIONAL
[R2-11]	Monitor submissions for frequent repetition of the same sample, therefore frequent detections would be seen, opening the chance of detecting outbreaks	OPTIONAL
[R2-12]	Monitor the submission system for statistical anomalies	OPTIONAL

Table R3. Firewalls

[R3-1]	Information acquired from firewalls	ESSENTIAL
[R3-2]	Include source name and location	ESSENTIAL
[R3-3]	Source IP of packet, protocol, src port, dst port, timestamp. Destination IP's may be anonymized. If anonymized, the system should indicate if there is a way to reverse the anonymization process and how.	ESSENTIAL

[R3-4]	Source IPs will have the following information attributed to them:		
	a.	Country location	DESIRABLE
	b.	ISP	DESIRABLE
	c.	Autonomous system number	DESIRABLE
[R3-5]	All the information will be available in near real-time		DESIRABLE
[R3-6]	Search capability will be available allowing for a search with IP or time information as key parameters		DESIRABLE

Table R4. IDS sensors

[R4-1]	Information acquired from IDS sensors		DESIRABLE
[R4-2]	Include source name and location		DESIRABLE
[R4-3]	If malware is found and reported it will allow for the possibility of supplying the information as specified in Malware Collections requirements, including:		
	a.	IPs contacted, if any (possibly used for C&C)	DESIRABLE
	b.	Exploit used if identified	DESIRABLE
[R4-4]	All the information will be available in near real-time.		DESIRABLE
[R4-5]	If the sensor is network-based:		
	a.	Source IP address of the logged event, protocol, src port, dst port, timestamp.	ESSENTIAL
	b.	A passive or active fingerprint identifying the OS system of the attacker will also be present.	DESIRABLE
	c.	Destination IPs will be anonymized	DESIRABLE
	d.	A description of the alert, in case of misuse-based sensor	ESSENTIAL
	e.	Description of an alert will be machine readable in some way, in case of misuse-based sensor	DESIRABLE
	f.	A threat level or threat probability, in case of anomaly-based sensor	ESSENTIAL
	g.	Packet content considered "unusual" will be supplied (potentially new exploits) in pcap format if possible along with flow information, in case of anomaly-based sensor	DESIRABLE
	h.	These packets will be screened so that their packet content does not disclose any private information, in case of anomaly-based sensor	DESIRABLE
	i.	If this is not possible, then such information does not need to be supplied (but it will be supplied), in case of anomaly-based sensor	OPTIONAL
	j.	Source IPs will have the country location information attributed to them, in case of anomaly-based sensor	DESIRABLE

[R4-6]	If the sensor is host-based:		
	a.	A description of the logged event	ESSENTIAL
	b.	Information related to the application or process which triggered it	DESIRABLE
	c.	A description of the target host	ESSENTIAL
	d.	A description of the alert, in case of misuse-based sensor	ESSENTIAL
	e.	Description of an alert will be machine readable in some way, in case of misuse-based sensor	DESIRABLE
	f.	A threat level or threat probability, in case of anomaly-based sensor	ESSENTIAL
	g.	Activity traces considered "unusual" will be supplied (potentially new exploits)	DESIRABLE
	h.	These activity traces will be screened so that their content does not disclose any private information, in case of anomaly-based sensor	DESIRABLE
	i.	If this is not possible, then such information does not need to be supplied (but it will be supplied), in case of anomaly-based sensor	OPTIONAL
	j.	Source IPs, if any obtained at this level, will have the country location information attributed to them, in case of anomaly-based sensor	DESIRABLE

Table R5. Honeypots

[R5-1]	Information acquired from Honeypots sensors		ESSENTIAL
[R5-2]	Source name and location		ESSENTIAL
[R5-3]	Source IP of the event (packet), protocol, src port, dst port, timestamp registered by the honeypot.		ESSENTIAL
[R5-4]	A fingerprint identifying the OS system of the attacker will also be present		DESIRABLE
[R5-5]	Destination IPs will be anonymized		DESIRABLE
[R5-6]	Source IPs will have the following information attributed to them:		
	a.	Country location	DESIRABLE
	b.	ISP	DESIRABLE
	c.	Autonomous system number	DESIRABLE
[R5-7]	Malware found information (if any), should comply with the Malware Collections requirements and additionally:		
	a.	IPs contacted, if any (possibly used for C&C)	DESIRABLE
	b.	Exploit used if identified	DESIRABLE
	c.	Optionally any traffic conversations recorded	DESIRABLE
[R5-8]	Packet content considered "unusual" will be supplied (potentially new exploits) in pcap format if possible along with flow information		DESIRABLE
[R5-9]	These packets will be screened so that their packet content does not disclose any private information. If this is not possible, then such information does not need to be supplied		DESIRABLE
[R5-10]	All the information will be available in near real-time		DESIRABLE
[R5-11]	Search capability will be available allowing for a search with IP information and timestamps as key parameters		DESIRABLE
[R5-12]	Any other types of data, if available, such as models used for detection of threats (e.g. ScriptGen), memory dumps, traces of exchanges, …		DESIRABLE

Table R6. Honeyclients

[R6-1]	Information acquired from honeyclients		ESSENTIAL
[R6-2]	Source name and location of the honeyclient		ESSENTIAL
[R6-3]	URL considered malicious or suspicious in case of Web served malware		ESSENTIAL
[R6-4]	The method by which URL was observed, such as web crawl with specific search parameters, spam URL, spim URL, user submission, other		ESSENTIAL
[R6-5]	Any associated exploit information if recognized (at least a name of the exploit if possible)		DESIRABLE
[R6-6]	Alert information pertaining to whether this URL successfully exploited latest patched versions of the OS		ESSENTIAL
[R6-7]	Malware found information will be in compliance with the Malware Collection requirements and additionally provide: IPs contacted, if any (possibly used for C&C)		DESIRABLE
[R6-8]	Associated URL information will be present		
	a.	Country location	DESIRABLE
	b.	ISP	DESIRABLE
	c.	Autonomous system number	DESIRABLE
	d.	List of IPs seen pointing to the site	DESIRABLE
	e.	First seen and last seen timestamp	DESIRABLE
	f.	Whois information associated with the URL	DESIRABLE
	g.	Whether suspected drive-by-download or not	OPTIONAL
	h.	Whether phishing URL or not	DESIRABLE
[R6-9]	All the information will be available in near real time		DESIRABLE
[R6-10]	Search capabilities will be available regarding data stored by the honeyclients, such as memory dumps, interaction traces, …		DESIRABLE

Table R7. Darknets

[R7-1]	Information acquired from darknets		DESIRABLE
[R7-2]	Source name and location		ESSENTIAL
[R7-3]	Source IPs of query, protocol, src port, dst port, timestamp. Destination IPs will be anonymized. This information will be available only "on demand", as darknet datasets are very large, through a search option		DESIRABLE
[R7-4]	Aggregated statistics will be available that show at least one of: amount of flows or packets or bytes in set periods for a darknet source		ESSENTIAL
[R7-5]	Associated IP information will be present:		
	a.	Country location	OPTIONAL
	b.	ISP provider	OPTIONAL
	c.	Autonomous system number	OPTIONAL
[R7-6]	All the information will be available in near real time		DESIRABLE
[R7-7]	Alerting information from a darknet source will be available, such as automated notifications about sharp increases in traffic for example or some specific alerts associated with a particular solution		OPTIONAL

Table R8. Mobile Devices

[R8-1]	Information acquired from mobile devices		DESIRABLE
[R8-2]	Source name and location		DESIRABLE
[R8-3]	For Bluetooth/WiFi/WiMAX-based sensors:		
	a.	The source address of the logged event (which should be anonymized, as it allows to track unique devices), service used, timestamp	DESIRABLE
	b.	A passive or active blueprinting of the attacking system will also be present	DESIRABLE
	c.	The sensor will supply captured data in a suitable format	DESIRABLE
	d.	If the service is a file-transfer service, the sensor will supply the transferred file	DESIRABLE

Table R9. RFID

[R9-1]	Information acquired from RFID	DESIRABLE
[R9-2]	Include data (tag ID, reader ID, tag data, timestamps) from large-scale RFID deployments	DESIRABLE
[R9-3]	Sharing of custom-written software modules	DESIRABLE

Table R10. Early Warning Systems

[R10-1]	Information acquired from Early Warning Systems		DESIRABLE
[R10-2]	Source name and location		DESIRABLE
[R10-3]	Early warning system information will be supplied to WOMBAT as soon as an EWS detects suspect activity		DESIRABLE
[R10-4]	Alerts produced by the system along with relevant associated information:		
	a.	Source IPs involved, protocol, src port, dst port, timestamp. Destination IPs will be anonymized	DESIRABLE
	b.	Payload information in pcap format if available	DESIRABLE
	c.	Any associated information specific to the EWS (such as snort alerts, EWS operator comments, threat signatures, detection models etc)	DESIRABLE
	d.	Clear criteria expressing why this alarm was generated and what it may mean	DESIRABLE
[R10-5]	Any IP information supplied will have the following associated with it:		
	a.	Country location	DESIRABLE
	b.	ISP provider	DESIRABLE
	c.	Autonomous system number	DESIRABLE
[R10-6]	Warnings prioritisation and confidence		DESIRABLE

4.1.4 Data Repository

The database of the WOMBAT system is intended to store all the relevant aggregated data as well as various types of metadata that will come out as a result of the data enrichment and characterization process. Requirements for the data repository have to address the way of storing the diverse data collected, as well as security and privacy considerations concerning data storage and data access. First issue will be satisfied by functional requirements that are listed in the table below. Second issue requires non-functional requirements which are specified in Section 5.5 and also in Section 6.1 that describes API design.

Table R11. Requirements for Data Repository

[R11-1]	WOMBAT will be able to share data provided by sensors (both existing and added in the future)	ESSENTIAL
[R11-2]	Data gathered from the sensors will be partially stored in the centralized way (central Data Repository). Remaining data will be kept at partners sites and will be accessible through a specially designed set of interfaces.	ESSENTIAL
[R11-3]	Database management system (e.g. mysql) will be set up to hold all the data and metadata collected	ESSENTIAL
[R11-4]	The system will be able to classify and store the heterogenic data from different sensors, that may vary greatly	ESSENTIAL
[R11-5]	The database schema will provide some extensibility to address both the heterogeneity of the data collected by the existing sensors and the fact that new forms of data may appear during the WOMBAT project	DESIRABLE
[R11-6]	The outgoing communication channel from database to client application or GUI will have one particular form (e.g. Web Service, HTTP)	DESIRABLE

4.2 Data Enrichment and Characterization

This section refers to the WP4 work-package of the WOMBAT project, particularly to WP4.1 (Code behavior), WP4.2 (Code structure), WP4.3 (Code context). The objective of WP4 is to develop techniques to characterize the malicious code collected in WP3, deriving from it metadata that might reveal insights into the origin of the code and the intentions of those that created, released or used it. The two main types of information are: (i) information about the actions of the code and its structure, (ii) and information about the context in which the code sample was collected. The following table specifies a set of requirements for WP4.

Table R12. Requirements for Data Enrichment and Characterization Process

[R12-1]	WOMBAT will provide a specification language (meaningful properties) to describe the behavior of machine-executable code (D 4.1)	ESSENTIAL
[R12-2]	WOMBAT will characterize the behavior of malicious code that is collected in database (D 4.2)	ESSENTIAL
[R12-3]	WOMBAT will provide characteristics of certain structures of malware code (i.e. PE structure, hashes of sections, entropy of that sections, CFG, etc.) (D 4.3 and D. 4.4)	ESSENTIAL

[R12-4]	WOMBAT will provide ways to use the properties ([R12-1]), together with contextual information to identify the miscreant behind malicious activity. The contextual information such as the country of origins of the attacks, timing, targets, etc. and results of [R12-2] and [R12-3] (D 4.5 and D 4.6)	ESSENTIAL
[R12-5]	WOMBAT will provide integration and correlation of different features used to describe malicious code, also with contextual information (D 4.7)	ESSENTIAL
[R12-6]	In the context of [R12-1], WOMBAT will provide new malware models (using e.g. grammars) to describe their behavior through their actions on the system. A model will be independent from the OS and the programming language used to create the malware.	DESIRABLE
[R12-7]	WOMBAT will allow static analysis of the malware's code (e.g. CFG analysis) with a controlled complexity (important, since this is a computationally intensive task) according to the available resources.	DESIRABLE
[R12-8]	WOMBAT will be able to identify and classify malware	ESSENTIAL
[R12-9]	WOMBAT will provide information from AV-engines about suspicious or malware binary files (how malware samples are detected by a list of antivirus vendors)	DESIRABLE
[R12-10]	WOMBAT will be able to identify certain malware samples related directly to online fraud	DESIRABLE
[R12-11]	WOMBAT will be able to provide URLs related to malware or online fraud	DESIRABLE
[R12-12]	WOMBAT will provide metadata, including: attack signatures, code behavior, structure of the malicious code	DESIRABLE
[R12-13]	WOMBAT will cluster together code exhibiting similar behavior	DESIRABLE
[R12-14]	Because of [R12-4], WOMBAT will provide information whether two code samples are related by origin	DESIRABLE
[R12-15]	WOMBAT will provide a phylogeny of code through static analysis of the binaries	DESIRABLE
[R12-16]	WOMBAT will create a model to infer behavior from code structure and phylogeny	DESIRABLE
[R12-17]	WOMBAT provide automatic inference of specification from binaries to the analysis of malware (e.g. by applying techniques of software engineering)	DESIRABLE
[R12-18]	WOMBAT will correlate and aggregate data from various sources and different types (binaries, firewall logs, honeynets data, etc.) (i.e. malware executables opens or connect to characteristic port will be correlate with statistics about network traffic to/from this port and countries of source/destination connections, etc.)	DESIRABLE

4.3 Threats Intelligence

This section refers to the work-package WP5 of the WOMBAT project which aims at understanding the root causes of the observed attacks, to better predict upcoming threats. This knowledge will form the basis for development of an early warning system. Requirements for the type of required analysis, models and techniques as well as the expected processing and analysis results are defined in the following table.

Table R13. Requirements for Threat Intelligence Process

[R13-1]	WOMBAT will be able to identify root causes of attacks by extracting the modus operandi of attackers from groups of related metadata using graph-based techniques and other data mining algorithms.	ESSENTIAL
[R13-2]	WOMBAT will be able to use and enhance models of normal malicious behavior to identify new emerging types of threats.	ESSENTIAL
[R13-3]	WOMBAT will use the clustering of seemingly unrelated data resulting from root cause analysis to detect stealthy malicious activities like multiheaded slow worms.	ESSENTIAL
[R13-4]	Assessing the quality of the results of all root cause analysis techniques implemented.	ESSENTIAL
[R13-5]	WOMBAT will include an Early Warning System (based on understanding of root causes of the attacks observed) to predict upcoming threats. The system will issue context-rich alerts, with references to similar activity in the past.	ESSENTIAL
[R13-6]	Assessing the quality of the results of WP4 by using EWS developed	ESSENTIAL
[R13-7]	WOMBAT will provide advanced search capabilities, that is, given some information about a piece of malware, it will be able to quickly query for related pieces (to do correlation).	ESSENTIAL
[R13-8]	WOMBAT will find patterns of related behavior.	DESIRABLE
[R13-9]	WOMBAT will identify shared code fragments between malware, indicative of common authorship.	DESIRABLE
[R13-10]	WOMBAT will identify origins of malware (where was it hosted, how are victims lured there).	DESIRABLE
[R13-11]	WOMBAT will enable real-time analysis along with a recording mechanism to restore the system after an attack.	DESIRABLE
[R13-12]	WOMBAT will search for general characteristics of different malware families, detecting patterns to protect against future mutations	OPTIONAL
[R13-13]	WOMBAT will allow to find crossing information regarding infrastructure used by malware.	DESIRABLE
[R13-14]	WOMBAT will use reincidence in the usage of certain infrastructure to identify malicious resource providers, or to warn non-malicious ones about the abuse of their infrastructure.	DESIRABLE
[R13-15]	WOMBAT will develop and use new malware models to evaluate the detection capabilities of the tools for detecting malware propagation	DESIRABLE

4.4 Data Output

This section specifies requirements for the results that are expected to come out from the threat data and information analysis within the WOMBAT project. In particular, they will include characteristics of threats. Results of analysis will be generated (amongst other information) from metadata stored in Data Repository (WP3) as well as from the output of Data Enrichment and Characterization (WP4) and Threats Intelligence (WP5) process. WOMBAT is intended to form the basis of a future worldwide early warning system, thus the potential result expected to be generated and distributed are: up-to-date information about new types of Internet security threats as well as ready-for-use attack and malware signature updates. The details of the information to be provided are listed in the following tables.

Table R14. Requirements for Threat Characteristics

	Information/ metadata generated in the process of Data Enrichment and Characterization (WP4)	
[R14-1]	Rankings and statistics of current and past malicious activity which will be divided (and correlated) by port number, type of system (HN, DN, etc.), country/ASN, vulnerability exploited, etc.	DESIRABLE
[R14-2]	Rankings and statistics of current and past malicious activity will be (thereafter) also distinguished by different data displayed on x-axis (to describe malicious behavior) such as: numbers of input or output flows, numbers of unique source or destination IP	DESIRABLE
[R14-3]	Information and statistics about attackers' OS (this information could be by packets analysis or malware binaries characterization)	OPTIONAL
[R14-4]	Information/metadata and statistics about origins of attacks, methods of attacks, all other results of characterization	DESIRABLE
[R14-5]	Information will be provided in a table format or plain text.	OPTIONAL
	Information generated in the process of Threats Intelligence (WP5)	
[R14-6]	Information about origins of malware	DESIRABLE
[R14-7]	Shared code fragments	DESIRABLE
[R14-8]	Models of malicious behavior and activity	DESIRABLE
[R14-9]	Results of assessment of malware and attacks' impact	DESIRABLE
[R14-10]	Results of all static and dynamic analysis	DESIRABLE
[R14-11]	Results of analysis of malware code presented in CFG (Control Flow Graph)	OPTIONAL

Table R15. Requirements for Early Warnings of Security Threats

[R15-1]	Early warnings will come from the system of alarms	DESIRABLE
[R15-2]	Alarms will be differentiated in terms of their kind and priority	ESSENTIAL
[R15-3]	Alarms will indicate detection of new threats or attacks, anomalies, increase of malicious activity	DESIRABLE
[R15-4]	Alarms will inform about new vulnerabilities	DESIRABLE

Table R16. Requirements for Virus and Attack Signatures Updates

[R16-1]	Meaningful (correlated and contextual) information and metadata	ESSENTIAL
[R16-2]	Descriptions of classes of malware related behavior	DESIRABLE
[R16-3]	Signatures of attacks	DESIRABLE
[R16-4]	Signatures will describe behavior on different levels, like network level (payloads of flows), or system/host level (memory dumps, system resources access, system registry in Windows, etc.)	OPTIONAL
[R16-5]	Signatures will be deliverable in different standards (to limit software, that could used this signatures)	OPTIONAL
[R16-6]	Universal models of malicious behavior and activity	DESIRABLE

Table R17. Requirements for Security Practices Updates

[R17-1]	Suggested security practices based on threat characteristics, warnings of security threats as well as (if generated) any clusters and signatures (example: block port X because of Y)	ESSENTIAL
[R17-2]	Security Practices will be deliverable via system Security Messages and reports (both periodic and instant), or other type of presentation and dissemination forms (news, articles, blog entries, RSS feed, notes and comments, newsletter via email)	DESIRABLE

5 NON-FUNCTIONAL REQUIREMENTS

This chapter specifies constraints on the WOMBAT system design and implementation, including: the kinds of operating systems that should be supported, issues concerned system integration, the set of its performance parameters. Since the system is intended to assure high interaction, on-line availability as well as sensitive and critical information protection, also requirements such as reliability, backup and recovery as well as security and privacy considerations, concerning information storage, processing and transfer, are carefully specified. Other non-functional requirements relate the ease of the system usage and specify sizing, scaling needs to meet planned growth.

5.1 System Environment

Table R18. Requirements for System Environment

[R18-1]	The core of the system will be based on Unix or a Unix-like operating system, the sources, including sensors may use any operating system based on requirements for a given task.	DESIRABLE
[R18-2]	The system's GUI will support standard or de-facto standard web components, and be accessible from all major operating systems and browsers, including at least Microsoft Windows, Linux, MacOS X, Internet Explorer, Firefox and Safari.	DESIRABLE
[R18-3]	The system will be modular.	DESIRABLE

5.2 Integration with Other Systems

Table R19. Requirements for Integration with Other Systems

[R19-1]	WOMBAT will include format conversion software (to convert data from other systems)	DESIRABLE
[R19-2]	WOMBAT will use XML as a preferred format for data exchange with external systems unless the system already offers an interface using another format.	OPTIONAL
[R19-3]	The confidentiality and integrity of communications will be specified whenever applicable.	DESIRABLE

5.3 System Performance

Table R20. Requirements for System Performance

[R20-1]	WOMBAT will support hundreds of simple queries per minute, where simple queries are defined as access to centrally stored data using typical, predefined queries.	DESIRABLE
[R20-2]	Response time for typical queries using only the central database will not exceed 10 seconds, response time for simple custom queries can be longer, but will not exceed 30 seconds.	DESIRABLE
[R20-3]	WOMBAT will support hundreds of simultaneous users.	DESIRABLE

[R20-4]	WOMBAT will support tens of custom analyses per minute.	OPTIONAL
[R20-5]	Complicated custom analyses will be performed in batch mode.	OPTIONAL

5.4 Reliability and Availability

Table R21. Requirements for System Environment

[R21-1]	The system will store raw data for at least a week and aggregated data for at least a month, the times may vary depending on the importance of data. Shorter data retention is possible as an exception only if necessary. Data retention will be adjusted according to legal and social requirements.	DESIRABLE
[R21-2]	The system will be robust – failures of individual sensors or even aggregated data sources may not cause a system failure.	ESSENTIAL
[R21-3]	The system's MTBF will be at least one month for failures repairable under one hour and three months for more serious failures.	DESIRABLE
[R21-4]	Allowable down time of the system will not exceed one full day per month.	DESIRABLE
[R21-5]	Routine activities related to system administration such as backup, user management and sources management will be performed without down time.	DESIRABLE
[R21-6]	In case of maintenance: (i) accepting of new queries will be stopped, (ii) the running queries will be completed, (iii) the down time will be announced at the web page.	ESSENTIAL

5.5 Security and Privacy

Table R22. Requirements for Security and Privacy

[R22-1]	The system will prevent access to personal data (including IP numbers) of targets and sources of attacks as well as of the information sources by regular, non-privileged users.	ESSENTIAL
[R22-2]	The system will protect malware samples form being accessed by non-privileged users.	ESSENTIAL
[R22-3]	User access to the database will be restricted to the API, such that user rights cannot be ignored.	ESSENTIAL
[R22-4]	The API will restrict access to data depending on user rights.	ESSENTIAL
[R22-5]	Access level for users will be managed by the WOMBAT consortium.	ESSENTIAL
[R22-6]	Access to data classified as publicly available (high-level statistics, etc.) will not require a decision by the consortium (“guest users”).	DESIRABLE

5.6 Usability

Table R23. Requirements for Usability

[R23-1]	The system will be highly interactive.	DESIRABLE
[R23-2]	The system will be convenient to use.	DESIRABLE
[R23-3]	Usability tests with potential users will be performed.	OPTIONAL
[R23-4]	The presentation of results will be clear.	DESIRABLE
[R23-5]	Two separate views will be available in the GUI – basic, for regular users seeking general information about threats and threat statistic, and expert, for advanced users performing tasks such as malware analysis.	DESIRABLE

5.7 Scalability

Table R24. Requirements for Scalability

[R24-1]	The system will be able to support data collection from at least 50 aggregated sources in the future, potentially reaching tens of thousands of individual sensors.	DESIRABLE
[R24-2]	The system will be able to store and process terabytes of raw data (tens of thousands of malware samples, millions of flows daily)	DESIRABLE
[R24-3]	Aggregation of data will be performed to avoid recomputing typical statistics on demand.	DESIRABLE
[R24-4]	The system will also be able to access raw data from individual sources without storing it locally	DESIRABLE

6 USER INTERFACE

WOMBAT is expected to collect a wealth of various data and information from heterogeneous systems (infrastructures and networks) and generate high quality results in the subsequent steps. This requires complete, friendly, helpful and effective user-oriented output for users. Such a user interface ESSENTIAL provide and aggregate all collected and generated KID (knowledge, information and data), statistics, results of different kinds of performed analysis and other results of Threat Intelligence. However, privacy aspects ESSENTIAL be considered with care. User interface ESSENTIAL support critical KID and privacy protection through using different dissemination level.

6.1 API Design

The API will be the lowest level interface, that provides support requests from outside of the WOMBAT system for all KID kept in database.

Table R25. Requirements for API Design

[R25-1]	API will be an intermediate system between data repository layer and output/clients	ESSENTIAL
[R25-2]	Access to the data will be governed by legal agreements singed between the consortium and the entity that requests access to the data. Depending on the time of agreement different entities will get different types of access.	ESSENTIAL
[R25-3]	Support secure connections between systems and secure data transfer	DESIRABLE
[R25-4]	Provide API-client software, that will communicate to API from client side	DESIRABLE
[R25-5]	API will use Web Services system (WSDL, SOAP, etc.) or compatible/similar system for future systems	OPTIONAL

6.2 Data Displaying and Graphical Visualisation

The Graphical User Interface (GUI) will provide project results in a user-friendly graphical layout.

Table R26. Requirements for GUI

[R26-1]	GUI will provide users with visual representations such as: Top IP addresses, top ports, attacks in the last hour, etc.	ESSENTIAL
[R26-2]	GUI will enable correlation through common attributes of different types of datasets and results of analysis This may be as simple as checking for IPs across different input systems, or more complex such as looking for similarities across different code.	DESIRABLE
[R26-3]	GUI will be based on the End-User-API (may communicate with system resources via End-User-API)	OPTIONAL
[R26-4]	GUI will be provided via web-page (web-page may communicate with Web-Services used by End-User-API)	OPTIONAL

[R26-5]	GUI will support dissemination levels and provide different user accounts		DESIRABLE
[R26-6]	GUI will sanitize information (IPs, host names, etc.). Level of sanitization ESSENTIAL depend on type of user account		DESIRABLE
[R26-7]	GUI will allow for system administration (only for superusers), and and will provide a useable interface to display project results (for users and superusers)		DESIRABLE
[R26-8]	GUI will be platform and application independent		DESIRABLE
[R26-9]	GUI will provide different type of KID visualisation:		DESIRABLE
	a.	User-interactive (zooming, selection, etc.) rankings of network activity, malware activity, etc. divided/grouped by different kinds of data (honeynets, darknets, viruses, etc.)	OPTIONAL
	b.	Tables with statistics of different kinds of data (flow stats, cluster stats, darknet stats, malware stats, etc.)	DESIRABLE
	c.	Binary and/or malware visualization solutions	DESIRABLE
	d.	Visualizations of system status (status of sensors, subsystems, hardware, etc.)	DESIRABLE
	e.	*Parallel coordinate plots* for different type of data	OPTIONAL
	f.	Graphs to describe malware and network activity behavior	OPTIONAL
	g.	*Alerts map* to visualize system alarms and system status if any	OPTIONAL
	h.	Auto-generated periodic reports	DESIRABLE
[R26-10]	GUI will provide different layouts: the most advanced - for computers (PCs) and limited - for mobile devices (smart phones, palmtops, etc.)		DESIRABLE
[R26-11]	GUI will provide community tools to help aggregate and share information, thoughts and ideas between members of consortium (power users), such as forums, blogs and wiki		DESIRABLE
[R26-12]	Community tools will be used also to provide information to public users via the official website		OPTIONAL

7 TESTING AND EVALUATION

The following table of testing requirements defines what ESSENTIAL/DESIRABLE/ OPTIONAL be checked during the WOMBAT system's testing procedures. System tests will cover both functional and non-functional requirements. Functional requirements are defined as a set of software deliverables to be proposed and implemented in work-packages WP3, WP4 and WP5. Non-functional requirements define quality of a developed solution by evaluation of the features defined in subsections 5.3-5.7.

Table R27. Requirements for the WOMBAT System Testing

	Success indicators of the system **functional requirements** design and implementation:		
[R27-1]	a.	Observation of a collective growth of over 20%* of malware samples during the first year of operation	ESSENTIAL
	b.	Observation of a collective growth of over 100%* of malware samples during 3 years of operation	ESSENTIAL
[R27-2]	a.	Qualification of each malware by at least 5 contextual information at the end of 2nd year of the project	ESSENTIAL
	b.	Qualification of each malware by at least 8 contextual information at the end of the project	ESSENTIAL
[R27-3]	a.	Specification of one threat intelligence model at the end of the 2nd year of the project	ESSENTIAL
	b.	Implementation of threat intelligence model at the end of the project	ESSENTIAL
[R27-4]	**Coverage** will be assessed using a defined methodology according to well-defined criteria. The methodology will consider different contexts (e.g. both known and unknown threats) and should remain independent from any editors/ organizations.		DESIRABLE
	Efficiency and scalability:		
[R27-5]	Efficiency and scalability will be analyzed in context of different data source and data source type numbers.		DESIRABLE
[R27-6]	Efficiency will be analyzed considering size of data stored in repositories proposed in the system architecture.		DESIRABLE
[R27-7]	Scalability will be analyzed considering number of concurrent users of the system.		DESIRABLE
[R27-8]	Requirements for network bandwidth allocation will be tested and estimated.		DESIRABLE
	Reliability:		
[R27-9]	Fault tolerance to particular system nodes failures will be analyzed.		DESIRABLE
[R27-10]	System immunity to potential side-effects of unknown malware analysis will be tested.		DESIRABLE
	Security and privacy:		

[R27-11]	It will be checked whether it is forced that only legitimate users may have access to data and algorithms made available to them due to some approved security policy.	DESIRABLE
[R27-12]	It will be checked whether there is no leak of private data to the central repository against privacy policies of the individual data repositories and their owners.	DESIRABLE
	Other:	
[R27-13]	The system will have document with proposed plan of tests.	DESIRABLE
[R27-14]	All results from the test procedures will be gathered in written test reports.	DESIRABLE
[R27-15]	During the project some procedure will be established for notification, discussion and feedback for error reports.	DESIRABLE
[R27-16]	System implementation will involve some software project management and bug-tracking software.	OPTIONAL
[R27-17]	All the modules will be tested separately before the final tests of the whole system.	DESIRABLE
[R27-18]	Each module will implement unit tests to enable semi-automatic testing of per module changes influence for the proper functioning of other modules behavior.	DESIRABLE

* percentage of additional malware collected related to the volume of collected malware in individual project particitpant's databases

8 CONFIGURATION MANAGEMENT

Configuration management defines management of the software during development and testing processes of the WOMBAT solution and management of the configuration of already completed software during its normal future exploitation. Requirements for the configuration management of the WOMBAT can be considered either as a requirements for a software project or for a deployed solution. Corresponding requirements are presented in the following table.

Table R28. Requirements for WOMBAT Testing

	For a software project:	
[R28-1]	Modules of the system developed separately by project participants will have separate private repositories	DESIRABLE
[R28-2]	Modules of the system developed separately by project participants will use some versioning system	OPTIONAL
[R28-3]	Modules of the system developed by participants as a common effort will have established shared repository with some software version control system	ESSENTIAL
[R28-4]	Apart from production deployment WOMBAT will be also installed on dedicated hardware provided by participants and configured to create the testing environment	ESSENTIAL
	For a deployed solution:	
[R28-5]	System will have single configuration stored in one place which describes: ▪ legitimate users with authentication information, ▪ run-time parameters, ▪ list of modules which constitute current configuration	DESIRABLE
[R28-6]	System will have flexible updating procedure for future versions of the system modules	OPTIONAL

APPENDIX A

The following table (Table R-U) is a reference, providing an overview of priorities and reasons for inclusion for all the system requirements. The second column validates the need for a given requirement and contains a list of reasons for inclusion. The reasons can be:

- FORMAL – the requirement is a direct consequence of the Description of Work document and is in fact one of the project's goals – possibly minor, but formally specified;
- INTERNAL – the requirement follows from best practices and is necessary for the project to reach completion, also used for basic requirements whose absence would make most of the user requirement impossible to achieve;
- [*user requirement number*] – the requirement is necessary to fulfill a user requirement (in many cases only the most important links are listed), can also be used with INTERNAL to signify that an especially strong relationship exists between this user requirement and the system requirements.

The third column shows the priority of the requirement (this information is also present in the main document and is repeated here for reference only).

Table R-U. The WOMBAT System and Users Requirements Associations

NO. OF SYSTEM REQUIREMENT	REASON FOR INCLUSION	REQUIREMENT PRIORITY
4. FUNCTIONAL AND DATA REQUIREMENTS		
4.1 Data Collection and Distribution		
[R1-1]	INTERNAL, [U1-4], [U2-3], [U5-2]	DESIRABLE
[R1-2]	INTERNAL, [U1-4], [U2-3], [U5-2]	DESIRABLE
[R1-3]	INTERNAL	DESIRABLE
[R1-4]	INTERNAL	DESIRABLE
[R1-5]	INTERNAL	DESIRABLE
[R1-6]	INTERNAL, [U1-4], [U2-3], [U5-2]	DESIRABLE [a,b,c,d]
[R1-7]	INTERNAL, [U1-4], [U2-3], [U5-2]	OPTIONAL
[R1-8]	INTERNAL, [U1-2], [U1-5], [U4-7], [U5-2], [U6-1]	DESIRABLE
[R1-9]	[U1-4], [U2-3]	OPTIONAL
[R1-10]	INTERNAL	DESIRABLE
[R1-11]	[a-e] FORMAL, [U1-2], [U1-4] [f] FORMAL, INTERNAL, [U1-2], [U1-4] [g] FORMAL, INTERNAL	ESSENTIAL [a-g]
[R2-1]	FORMAL, [U1-1], [U1-2], [U1-5], [U4-1], [U5-5], [U6-1]	ESSENTIAL
[R2-2]	INTERNAL	ESSENTIAL
[R2-3]	[U1-2], [U1-5], [U4-1], [U6-1]	ESSENTIAL
[R2-4]	[U1-2], [U1-5], [U4-1], [U6-1]	DESIRABLE
[R2-5]	[U5-2]	ESSENTIAL
[R2-6]	[U1-2], [U1-5], [U4-1], [U6-1]	DESIRABLE
[R2-7]	[U1-2], [U1-5], [U4-1], [U6-1]	DESIRABLE
[R2-8]	[U1-2], [U1-5], [U4-1], [U5-2], [U6-1]	DESIRABLE
[R2-9]	[U1-2], [U1-5], [U4-1], [U6-1]	DESIRABLE
[R2-10]	[U1-2], [U1-5], [U4-1], [U6-1]	OPTIONAL
[R2-11]	[U1-2], [U1-5], [U4-1], [U6-1]	OPTIONAL
[R2-12]	[U1-3], [U2-5], [U3-1], [U3-3], [U3-5], [U4-7], [U5-6], [U5-7], [U7-2]	OPTIONAL

[R3-1]	FORMAL, [U2-2], [U2-4], [U3-1], [U3-3], [U5-5], [U6-3], [U7-1]	ESSENTIAL
[R3-2]	[U5-2], [U6-1], [U3-1]	ESSENTIAL
[R3-3]	[U2-4], [U3-1], [U6-3]	ESSENTIAL
[R3-4]	[U2-4], [U3-1], [U6-3], [U5-6]	DESIRABLE [a,b,c]
[R3-5]	[U3-5], [U4-7], [U5-7]	DESIRABLE
[R3-6]	[U3-1], [U3-3], [U6-3]	DESIRABLE
[R4-1]	FORMAL, [U2-2], [U3-1], [U3-5], [U5-5], [U6-3], [U7-1]	DESIRABLE
[R4-2]	[U3-1], [U5-2], [U6-1]	DESIRABLE
[R4-3]	[U1-2], [U1-3], [U1-4], [U1-5], [U2-1], [U2-2], [U3-1], [U3-6], [U4-1], [U4-3], [U5-3], [U5-5], [U5-7], [U6-3], [U7-2]	DESIRABLE [a,b]
[R4-4]	[U3-5], [U4-7], [U5-7]	DESIRABLE
[R4-5]	[U1-2], [U1-7], [U2-4], [U3-1], [U4-7], [U5-1], [U5-2], [U5-6], [U5-7], [U6-3]	ESSENTIAL [a,d,f] DESIRABLE [b,c,e,g,h,j] OPTIONAL [i]
[R4-6]	[U1-2], [U1-7], [U2-4], [U3-1], [U4-7], [U5-1], [U5-2], [U5-6], [U5-7], [U6-3]	ESSENTIAL [a,c,d,f] DESIRABLE [b,e,g,h,j] OPTIONAL [i]
[R5-1]	FORMAL, [U2-2], [U2-4], [U3-1], [U3-3], [U5-5], [U6-3], [U7-1]	ESSENTIAL
[R5-2]	[U3-1], [U5-2], [U6-1]	ESSENTIAL
[R5-3]	[U3-1], [U6-3]	ESSENTIAL
[R5-4]	[U3-5], [U5-7]	DESIRABLE
[R5-5]	INTERNAL	DESIRABLE
[R5-6]	[U2-4], [U3-1], [U6-3], [U5-6]	DESIRABLE [a,b,c]
[R5-7]a-c	[U1-2], [U1-3], [U1-4], [U1-5], [U2-1], [U2-2], [U3-1], [U3-6], [U4-1], [U4-3], [U5-3], [U5-5], [U5-7], [U6-3], [U7-2]	DESIRABLE [a,b,c]
[R5-8]	[U1-4], [U3-5], [U4-7], [U5-7]	DESIRABLE
[R5-9]	INTERNAL	DESIRABLE
[R5-10]	[U3-5], [U4-7], [U5-7]	DESIRABLE
[R5-11]	[U3-1], [U3-3], [U6-3]	DESIRABLE
[R5-12]	[U1-3], [U2-2], [U3-3], [U4-7], [U5-7], [U6-1]	DESIRABLE
[R6-1]	FORMAL, [U2-2], [U2-4], [U3-1], [U3-3], [U4-2], [U4-5], [U5-4], [U5-5], [U5-6], [U6-2], [U6-3], [U7-1]	ESSENTIAL
[R6-2]	[U3-1], [U5-2], [U6-1]	ESSENTIAL
[R6-3]	[U3-1], [U4-2], [U4-5], [U5-4], [U6-2]	ESSENTIAL

[R6-4]	[U5-2]	ESSENTIAL
[R6-5]	[U7-2]	DESIRABLE
[R6-6]	[U1-4], [U3-5], [U4-7], [U5-6], [U5-7], [U7-2]	ESSENTIAL
[R6-7]	[U3-1], [U6-3]	DESIRABLE
[R6-8]	[U1-2], [U1-5], [U3-1], [U4-3], [U5-3], [U6-1]	DESIRABLE [a,b,c,d,e,f,h] OPTIONAL [g]
[R6-9]	[U3-5], [U4-7], [U5-7]	DESIRABLE
[R6-10]	[U1-5], [U3-1], [U6-3]	DESIRABLE
[R7-1]	FORMAL, [U2-2], [U2-4], [U3-1], [U3-3], [U5-5], [U6-3], [U7-1]	DESIRABLE
[R7-2]	[U3-1], [U5-2], [U6-1]	ESSENTIAL
[R7-3]	[U3-1], [U6-3]	DESIRABLE
[R7-4]	[U2-4], [U3-1], [U6-3]	ESSENTIAL
[R7-5]	[U2-4], [U3-1], [U6-3], [U5-6], [U7-1]	OPTIONAL [a,b,c]
[R7-6]	[U3-5], [U4-7], [U5-7]	DESIRABLE
[R7-7]	[U3-5], [U4-7], [U5-6], [U5-7], [U7-1]	OPTIONAL
[R8-1]	FORMAL, [U5-6], [U4-7], [U5-7]	DESIRABLE
[R8-2]	[U5-2]	DESIRABLE
[R8-3]	[U1-4], [U3-5], [U4-7], [U5-7], [U6-1]	DESIRABLE [a,b,c,d]
[R9-1]	FORMAL, [U5-6], [U4-7], [U5-7]	DESIRABLE
[R9-2]	FORMAL, [U5-6], [U4-7], [U5-7]	DESIRABLE
[R9-3]	[U2-1], [U6-1]	DESIRABLE
[R10-1]	FORMAL, [U2-2], [U2-4], [U3-1], [U3-3], [U4-2], [U4-5], [U5-4], [U5-5], [U5-6], [U6-2], [U6-3], [U7-1]	DESIRABLE
[R10-2]	[U3-1], [U5-2], [U6-1]	DESIRABLE
[R10-3]	[U3-5], [U4-7], [U5-7]	DESIRABLE
[R10-4]	[U2-4], [U3-1], [U3-3], [U3-5], [U4-1], [U4-7], [U5-1], [U5-7], [U7-2]	DESIRABLE [a,b,c,d]
[R10-5]	[U2-5], [U3-1], [U6-3], [U5-6]	DESIRABLE [a,b,c]
[R10-6]	INTERNAL, [U5-2]	DESIRABLE
[R11-1]	FORMAL, INTERNAL	ESSENTIAL
[R11-2]	INTERNAL	ESSENTIAL
[R11-3]	INTERNAL	ESSENTIAL
[R11-4]	FORMAL, INTERNAL	ESSENTIAL
[R11-5]	INTERNAL	DESIRABLE
[R11-6]	INTERNAL	DESIRABLE

4.2 Data Enrichment and Characterization		
[R12-1]	FORMAL	ESSENTIAL
[R12-2]	FORMAL	ESSENTIAL
[R12-3]	FORMAL	ESSENTIAL
[R12-4]	FORMAL	ESSENTIAL
[R12-5]	FORMAL	ESSENTIAL
[R12-6]	FORMAL, [U1-2], [U2-2], [U2-7], [U3-2]	DESIRABLE
[R12-7]	[U1-2], [U1-5]	DESIRABLE
[R12-8]	INTERNAL, [U1-1], [U1-2], [U1-3], [U1-5], [U6-1]	ESSENTIAL
[R12-9]	[U1-2], [U1-5], [U2-1], [U4-7], [U5-6], [U5-7], [U6-1]	DESIRABLE
[R12-10]	[U4-2], [U4-3], [U5-4], [U5-6]	DESIRABLE
[R12-11]	[U1-5], [U4-2], [U4-3], [U5-4], [U5-6]	DESIRABLE
[R12-12]	[U1-2], [U1-3], [U1-5], [U4-4], [U6-1]	DESIRABLE
[R12-13]	FORMAL, [U1-6], [U3-4], [U3-6], [U4-3], [U5-6], [U6-1]	DESIRABLE
[R12-14]	[U1-6], [U3-4], [U3-6], [U4-3], [U6-1]	DESIRABLE
[R12-15]	FORMAL, [U1-6], [U3-4], [U3-6], [U4-3], [U6-1]	DESIRABLE
[R12-16]	FORMAL, [U1-6], [U3-4], [U3-6], [U4-3], [U6-1]	DESIRABLE
[R12-17]	FORMAL, [U1-6], [U3-4], [U3-6], [U4-3], [U6-1]	DESIRABLE
[R12-18]	FORMAL, [U1-6], [U3-4], [U3-6], [U4-3], [U6-1]	DESIRABLE
4.3 Threat Intelligence		
[R13-1]	FORMAL	ESSENTIAL
[R13-2]	FORMAL	ESSENTIAL
[R13-3]	FORMAL	ESSENTIAL
[R13-4]	FORMAL	ESSENTIAL
[R13-5]	FORMAL	ESSENTIAL
[R13-6]	FORMAL	ESSENTIAL
[R13-7]	[U1-5], [U1-6], [U3-4], [U3-6], [U4-3], [U6-1]	ESSENTIAL
[R13-8]	[U1-5], [U1-6], [U3-4], [U3-6], [U4-3], [U6-1]	DESIRABLE
[R13-9]	[U1-5], [U1-6], [U3-4], [U3-6], [U4-3], [U6-1]	DESIRABLE

[R13-10]	[U1-5], [U1-6], [U3-1], [U3,4], [U3-6], [U4-1], [U4-3], [U4-5], [U6-1], [U6-3]	DESIRABLE
[R13-11]	INTERNAL	DESIRABLE
[R13-12]	[U1-5], [U1-6], [U3-4], [U3-6], [U4-3], [U6-1]	OPTIONAL
[R13-13]	[U1-5], [U1-6], [U3-4], [U3-6], [U4-3], [U6-1]	DESIRABLE
[R13-14]	[U1-5], [U1-6], [U3-1], [U3-4], [U3-6], [U4-1], [U4-3], [U4-5], [U6-1], [U6-3]	DESIRABLE
[R13-15]	[U1-5], [U1-6], [U3-5], [U3-6], [U4-7], [U5-6], [U5-7], [U6-1]	DESIRABLE
4.4 Data Output		
[R14-1]	[U2-1], [U2-4], [U3-3], [U4-7], [U5-7], [U7-1], [U7-2]	DESIRABLE
[R14-2]	[U2-1], [U2-4], [U3-3], [U4-7], [U5-7], [U7-1], [U7-2]	DESIRABLE
[R14-3]	[U2-1], [U2-4], [U3-3], [U4-7], [U5-7], [U7-1], [U7-2]	OPTIONAL
[R14-4]	[U2-1], [U2-4], [U3-3], [U4-7], [U5-7], [U7-1], [U7-2]	DESIRABLE
[R14-5]	INTERNAL	OPTIONAL
[R14-6]	[U1-6], [U3-6], [U4-3]	DESIRABLE
[R14-7]	[U1-6], [U3-6], [U4-3]	DESIRABLE
[R14-8]	[U1-6], [U3-6], [U4-3]	DESIRABLE
[R14-9]	[U1-6], [U3-6], [U4-3]	DESIRABLE
[R14-10]	[U1-6], [U3-6], [U4-3]	DESIRABLE
[R14-11]	[U1-6], [U3-6], [U4-3]	OPTIONAL
[R15-1]	INTERNAL	DESIRABLE
[R15-2]	INTERNAL	ESSENTIAL
[R15-3]	[U3-5], [U4-7], [U5-7]	DESIRABLE
[R15-4]	[U3-5], [U4-7], [U5-7]	DESIRABLE
[R16-1]	[U1-2], [U1-3], [U2-1], [U2-2], [U3-2], [U4-4], [U7-3], [U7-4], [U7-5]	ESSENTIAL
[R16-2]	[U1-2], [U1-3], [U2-1], [U2-2], [U3-2], [U4-4], [U7-3], [U7-4], [U7-5]	DESIRABLE
[R16-3]	[U1-2], [U1-3], [U2-1], [U2-2], [U3-2], [U4-4], [U7-3], [U7-4], [U7-5]	DESIRABLE
[R16-4]	[U1-2], [U1-3], [U2-1], [U2-2], [U3-2], [U4-4], [U7-3], [U7-4], [U7-5]	OPTIONAL
[R16-5]	[U1-2], [U1-3], [U2-1], [U2-2], [U3-2], [U4-4], [U7-3], [U7-4], [U7-5]	OPTIONAL

[R16-6]	[U1-2], [U1-3], [U2-1], [U2-2], [U3-2], [U4-4], [U7-3], [U7-4], [U7-5]	DESIRABLE
[R17-1]	[U2-1]	ESSENTIAL
[R17-2]	[U2-1]	DESIRABLE
5. NON-FUNCTIONAL REQUIREMENTS		
[R18-1]	INTERNAL	DESIRABLE
[R18-2]	INTERNAL	DESIRABLE
[R18-3]	INTERNAL	DESIRABLE
[R19-1]	INTERNAL	DESIRABLE
[R19-2]	INTERNAL	OPTIONAL
[R19-3]	INTERNAL	DESIRABLE
[R20-1]	INTERNAL	DESIRABLE
[R20-2]	INTERNAL	DESIRABLE
[R20-3]	INTERNAL	DESIRABLE
[R20-4]	INTERNAL	OPTIONAL
[R20-5]	INTERNAL	OPTIONAL
[R21-1]	INTERNAL	DESIRABLE
[R21-2]	INTERNAL	ESSENTIAL
[R21-3]	INTERNAL	DESIRABLE
[R21-4]	INTERNAL	DESIRABLE
[R21-5]	INTERNAL	DESIRABLE
[R21-6]	INTERNAL	ESSENTIAL
[R22-1]	INTERNAL	ESSENTIAL
[R22-2]	INTERNAL	ESSENTIAL
[R22-3]	INTERNAL	ESSENTIAL
[R22-4]	INTERNAL	ESSENTIAL
[R22-5]	INTERNAL	ESSENTIAL
[R22-6]	INTERNAL, [U7-1], [U7-2], [U7-3], [U7-4], [U7-5]	DESIRABLE
[R23-1]	INTERNAL	DESIRABLE
[R23-2]	INTERNAL	DESIRABLE
[R23-3]	INTERNAL	OPTIONAL
[R23-4]	INTERNAL	DESIRABLE
[R23-5]	INTERNAL	DESIRABLE
[R24-1]	FORMAL, INTERNAL	DESIRABLE

[R24-2]	INTERNAL	DESIRABLE
[R24-3]	INTERNAL	DESIRABLE
[R24-4]	INTERNAL	DESIRABLE
6. USER INTERFACE		
6.1 API Design		
[R25-1]	INTERNAL	ESSENTIAL
[R25-2]	FORMAL, INTERNAL	ESSENTIAL
[R25-3]	INTERNAL	DESIRABLE
[R25-4]	INTERNAL	DESIRABLE
[R25-5]	INTERNAL	OPTIONAL
6.2 Data Displaying and Graphical Visualization		
[R26-1]	INTERNAL	ESSENTIAL
[R26-2]	INTERNAL	DESIRABLE
[R26-3]	INTERNAL	OPTIONAL
[R26-4]	INTERNAL	OPTIONAL
[R26-5]	INTERNAL	DESIRABLE
[R26-6]	INTERNAL	DESIRABLE
[R26-7]	INTERNAL	DESIRABLE
[R26-8]	INTERNAL	DESIRABLE
[R26-9]	INTERNAL	DESIRABLE [b,c,d,h] OPTIONAL [a,e,f,g]
[R26-10]	INTERNAL	DESIRABLE
[R26-11]	INTERNAL	DESIRABLE
[R26-12]	INTERNAL	OPTIONAL
7. TESTING AND EVALUATION		
[R27-1]	FORMAL	ESSENTIAL [a,b]
[R27-2]	FORMAL	ESSENTIAL [a,b]
[R27-3]	FORMAL	ESSENTIAL [a,b]
[R27-4]	INTERNAL	DESIRABLE
[R27-5]	INTERNAL	DESIRABLE
[R27-6]	INTERNAL	DESIRABLE
[R27-7]	INTERNAL	DESIRABLE
[R27-8]	INTERNAL	DESIRABLE
[R27-9]	INTERNAL	DESIRABLE
[R27-10]	INTERNAL	DESIRABLE

[R27-11]	INTERNAL	DESIRABLE
[R27-12]	INTERNAL	DESIRABLE
[R27-13]	INTERNAL	DESIRABLE
[R27-14]	INTERNAL	DESIRABLE
[R27-15]	INTERNAL	DESIRABLE
[R27-16]	INTERNAL	OPTIONAL
[R27-17]	INTERNAL	DESIRABLE
[R27-18]	INTERNAL	DESIRABLE
8. CONFIGURATION MANAGEMENT		
[R28-1]	INTERNAL	DESIRABLE
[R28-2]	INTERNAL	OPTIONAL
[R28-3]	INTERNAL	ESSENTIAL
[R28-4]	INTERNAL	ESSENTIAL
[R28-5]	INTERNAL	DESIRABLE
[R28-6]	INTERNAL	OPTIONAL

REFERENCES

[1] Anubis (Analyzing Unknown Binaries), http://analysis.seclab.tuwien.ac.at/features.php

[2] ARAKIS, http://arakis.cert.pl/en/index.html

[3] Argos, http://www.few.vu.nl/argos/

[4] DeepSight Early Warning Services, Symantec, https://tms.symantec.com/Default.aspx

[5] DoW, WOMBAT - Description of Work

[6] Honey@Home, http://www.honeyathome.org/

[7] G. Jacob, E. Filiol and H. Debar: *Functional Polymorphic Engines: Formalisation, Implementation and Use Cases*, In Proceedings of the EICAR conference (Coming version in Journal of Computer Virology), 2008.

[8] Leurrecom.org Honeypot Project, Institut Eurécom, http://www.leurrecom.org/

[9] Virustotal, http://www.virustotal.com/