



# Deliverable 2.1

Yearly Updated Requirements Specification Documents

Symantec Ireland

Version 04

23.October 2006



IST 2006 027600

AntiPhish

Anticipatory Learning for Reliable Phishing Prevention

Specific Targeted Research or Innovation Project

2.4.3 Towards a global dependability and security framework

## D 2.1 Yearly Updated Requirements Specification Documents

Due date of deliverable: M09 (30 September 2006)

Actual submission date: 27 October 2006

Start date of project: 01. Jan. 2006

Duration: 36 months

Lead Contractor for this Deliverable: Symantec Ireland

Revision: 06

<b>Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)</b>		
<b>Dissemination Level</b>		
<b>PU</b>	Public	X
<b>PP</b>	Restricted to other programme participants (including the Commission Services)	
<b>RE</b>	Restricted to a group specified by the consortium (including the Commission Services)	
<b>CO</b>	Confidential, only for members of the consortium (including the Commission Services)	

## Revision history

Deliverable administration and summary		
Project acronym: AntiPhish		ID: IST-2006-027600
Document identifier:	AntiPhish-del-D21-RequirementsSpecification-f-v06	
Leading partner: Symantec Ireland		
Report version: v06		
Report preparation date: 29.Sep.2006		
Classification: Public		
Nature: Report		
Author(s) and contributors: Brian Witten and Patrick Horkan in collaboration with all partners		
Status:		Plan
		Draft
		Working
	X	Final
		Submitted
		Approved

The AntiPhish © Consortium has addressed all comments received, making changes as necessary. Changes to this document are detailed in the change log table below.

Date	Edited by	Status	Changes made
-	DoW	Plan	report template
06.10.06	Brian Witten	Working	First Draft
10.10.06	Brian Witten	Working	Second Draft
18.10.06	Brian Witten	Working	Incorporating Fraunhofer Suggestions
20.10.06	Brian Witten	Working	Incorporating additional suggestions
23.10.06	Brian Witten	Final	Corrected typographical errors
27.10.06	Gerhard Paaß	Submitted	Final consistency checks and preparation for submission to EC

Notice that other documents may supersede this document. A list of latest public AntiPhish deliverables can be found at the AntiPhish webpage at [www.AntiPhishResearch.org/publications](http://www.AntiPhishResearch.org/publications).

## Copyright

This report is © AntiPhish Consortium 2006. Its duplication is restricted to the personal use within the consortium, funding agency and project reviewers.

## Citation

Brian Witten (2006). Deliverable D2.1: 2006 Requirements Specification Document. AntiPhish Consortium, c/o Fraunhofer IAIS , [www.antiphishresearch.org](http://www.antiphishresearch.org)

## Acknowledgements

The work presented in this document has been conducted in the context of the EU Framework Programme project IST 2006 027600 AntiPhish. AntiPhish is a 36-month project that started on January 1st, 2006 and is funded by the European Commission as well as by the industrial partners. Their support is appreciated.

The partners in the project are Fraunhofer Institute for Intelligent Analysis and Information Systems (FHG), Symantec LIRIC Limited (LIRIC), Symantec Ltd. (Symantec Ireland), TISCALI Services S.r.l. (Tiscali) and K. U. Leuven / ICRI-LIIR (K.U. Leuven). The content of this document is the result of extensive discussions within the AntiPhish© Consortium as a whole.

## More information

Public AntiPhish reports and other information pertaining to the project are available through AntiPhish public web site under [www.antiphishresearch.org](http://www.antiphishresearch.org).

## Table of contents

Executive summary.....	6
1 Introduction .....	7
2 Critical Performance Parameters .....	8
3 Architectural Context .....	9
4 Requirements .....	10

## Executive summary

This document attempts to capture requirements for an improved system of blocking spam and phishing messages. Requirements are given for **prototype** and projected **production** implementations of policy enforcement points and an **analysis system** for extracting blocking criteria from labelled and unlabelled sources. Although other requirements are given in Section 3, Requirements, requirements for the prototype analysis system include:

Phishing False Negative Rate:	0.1
Phishing False Positive Rate:	0.01
Labelled Content Processing Message Throughput:	12 per second
Latency:	3,000 seconds
Personnel:	1 person

## 1 Introduction

This document attempts to capture requirements for an improved system of blocking spam and phishing messages. Requirements are given in context of **current** production system performance and required performance of **prototype** and projected **production** systems, along with projected **goals** for a production system. To provide greater context, we include an architectural depiction of the envisioned system.

This document specifically attempts to capture requirements for the portions of the system known as the **analysis system** which extracts blocking criteria from labelled and unlabelled sources, and disseminating those blocking criteria to **policy enforcement points**. This document additionally provides expected constraints on policy enforcement points for at least three reasons. First, the AntiPhish effort will build or leverage technologies for instantiating a limited number of policy enforcement points in test fielding. Second, constraints on policy enforcement points may impart constraints of the scale and complexity of blocking criteria that may be expressed by the analysis system to the policy enforcement points. Third, although substantial improvements to policy enforcement points may be possible through separate efforts, the overall commercial viability of the total system is sensitive to the rough costs of each policy enforcement point, and the scale of hardware for processing the volumes of information at policy enforcement points can vary dramatically if not intentionally bounded.

## 2 Critical Performance Parameters

The most critical performance parameters are described in Section 2.2 of the Technical Annex of the Commission Contract, and included again here in greater detail.

**Phishing False Negative Rate (PFN)** - The number of phishing messages not blocked at policy enforcement points divided by the number of phishing messages sent through policy enforcement points.

**Phishing False Positive Rate (PFP)** - The number of legitimate messages blocked (as phishing) at policy enforcement points divided by the number of total messages sent through policy enforcement points.

**Recall (RECALL)** – The ratio of phishing messages blocked divided by the number of total phishing messages.

**Precision (PRECISION)** – The ratio of phishing messages blocked divided by the total number of messages blocked.

**Labelled Content Processing Message Throughput (LCPMT)** – The volume of labelled data to be processed by the analysis system in generating blocking criteria to be disseminated to enforcement points, measured by the number of messages per second through the analysis system.

**Labelled Content Processing Volume Throughput (LCPVT)** – The volume of labelled data to be processed by the analysis system in generating blocking criteria to be disseminated to enforcement points, measured megabytes per second through the analysis system.

**Latency (LATENCY)** – The latency from the time a sample of a phishing is acquired from a labelled source to the time an effective rule can be deployed for blocking the phishing messages, measured in seconds.

**Personnel (PERSONNEL)** – The number of people required to operate the system at any given time to maintain the desired performance.

**Hardware Cost of the Analysis System (HCAS)** – The cost of hardware for the analysis system generating blocking criteria, measured in Euros.

**Hardware Cost Per Enforcement Point (HCPEP)** – The cost of hardware for enforcing blocking criteria at each enforcement point, measured in Euros.

**Projected Enforcement Point Message Throughput (PEPMT)** – The projected volume anticipated for each enforcement point, measured in messages per second. Please note that large customers may operate multiple policy enforcement points in parallel.

**Projected Enforcement Point Volume Throughput (PEPVT)** – The projected volume anticipated for each enforcement point, measured in megabytes per second. Please note that large customers may operate multiple policy enforcement points in parallel.

### 3 Architectural Context

The graphic below is provided for architectural context.

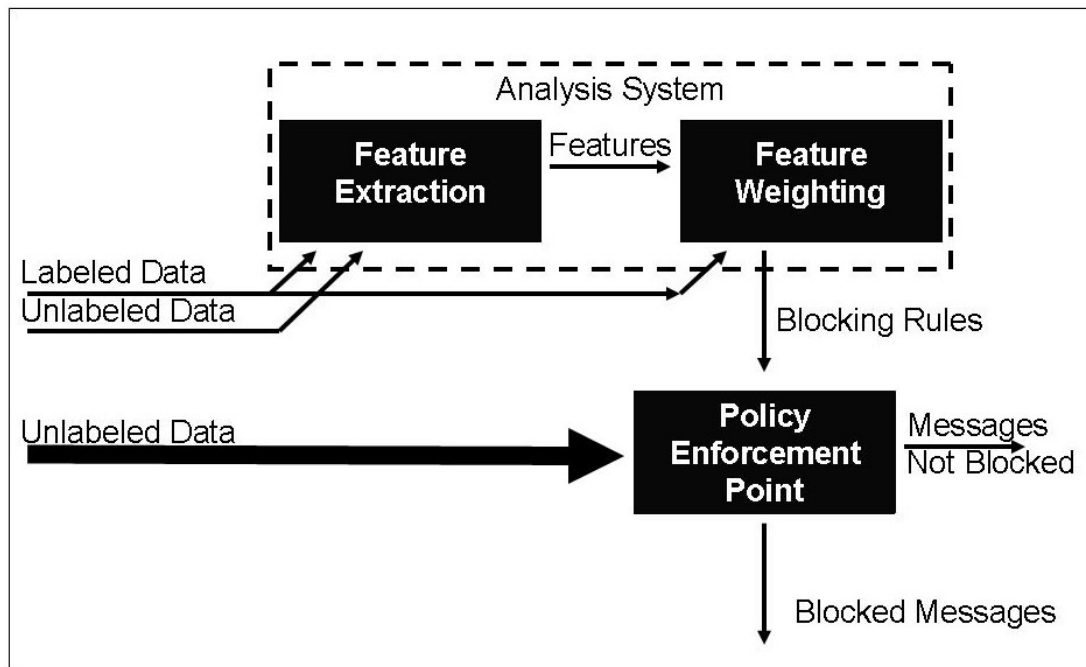


Figure 1: Architectural Depiction

This graphic depicts the relationships between the analysis system and the policy enforcement points, the relationships between labelled data and unlabeled data, and also the relationships between feature extraction and feature weighting.

Please note that automated feature weighting can already produce excellent performance in terms of recall and precision, when manual feature selection is possible. However, given that manual feature selection is not practical for the volume of phishing messages and the rate at which they change, automated feature selection is increasingly important. In this context, the KU Leuven research in feature selection and Fraunhofer machine learning based optimization of the analysis system as a whole may be exceptionally valuable.

Please also note that requirements for the interfaces between the analysis system and policy enforcement points are not constrained, and please note that the interface between feature extraction and feature weighting are also not constrained at this time since those interfaces should evolve on a path guided by results of the research currently being conducted by Fraunhofer and KU Leuven.

## 4 Requirements

	Current	Prototype	Production	Goal
<b>PFN</b>	proprietary	0.10*	0.05	0.01
<b>PEP</b>	proprietary	0.01	0.001	1/1M*
<b>RECALL</b>	proprietary	0.90	0.95	0.99
<b>PRECISION</b>	proprietary	0.95	0.99	0.999999
<b>LCPMT</b>	proprietary	12	1,200	10,000
<b>LCPVT</b>	proprietary	0.12	12	100
<b>LATENCY</b>	300	3,000	300*	30
<b>PERSONNEL</b>	proprietary	1	20	5
<b>HCAS</b>	proprietary	50,000	10M	< 250K
<b>HCPEP</b>	proprietary	10k	1M	10k
<b>PEPMT</b>	proprietary	100,000	100,000	1M
<b>PEPVT</b>	1,000	1,000	1,000	10,000

\* Note: These requirements were proposed for the Prototype in the original proposal to the European Commission.