



Deliverable 2.2

Yearly Updated Requirements Specification Document

Symantec Ireland

Version 07

15.December 2007



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 2.2 Yearly Updated Requirements Specification Document

Due date of deliverable: M23 (30 November 2007)

Actual submission date: 04 January 2008

Start date of project: 01. Jan. 2006

Duration: 36 months

Lead Contractor for this Deliverable: Symantec Ireland

Revision: 07

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

Revision history

Deliverable administration and summary		
Project acronym: AntiPhish		ID: IST-2006-027600
Document identifier:	AntiPhish-del-D22-RequirementsSpecification-f- v07	
Leading partner: Symantec Ireland		
Report version: v07		
Report preparation date: 30.Nov.2007		
Classification: Public		
Nature: Report		
Author(s) and contributors: Brian Witten and Patrick Horkan in collaboration with 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
17.10.07	Brian Witten	DRAFT	Update Reflecting a Year of Progress.
14.12.07	Brian Witten	DRAFT	Revised for comments on OCT draft.
04.01.07	Gerhard Paaß	Final	Final edits for submission

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.

Copyright

This report is © DataMiningGrid Consortium 2006. Its duplication is allowed only in the integral form for anyone's personal use for the purposes of research or education.

Citation

Brian Witten (2006). Deliverable D2.1: 2007 Requirements Specification Document. AntiPhish Consortium, c/o Fraunhofer IAIS , 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.



Table of contents

Executive summary	6
1 Introduction	8
2 Critical Performance Parameters	9
3 Architectural Context	11
4 Requirements	12



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 for the prototype analysis system include:

	Requirement	Current Prototype Performance
Phishing False Negative Rate:	Under 0.100	0.036
Phishing False Positive Rate:	Under 0.010	0.010
Latency:	3,000 seconds	TBD
Personnel:	1 person	1 person

We note that the latency performance remains “yet to be measured” as this depends on interactions between the centralized Analysis System and the distributed set of Policy Enforcement Points. Such interaction of the distributed system remains to be measured in the coming year. However, the system seems to be on-track for meeting the latency goal as the processing time seems to be under 10 seconds at both the Analysis System and Policy Enforcement Point.

Given the sensitivity of phishing detection to changes in effectiveness of feature extraction, given the sensitivity of effectiveness in feature extraction to previously unseen salting techniques, and from experience over the past year detecting emergence of new phishing techniques that employ new means of salting and mimicry to evade feature extraction, we also state more specific goals in this area. Specifically, these goals are a **Probability of False Positive in Detection of New Salting Techniques** under 1% and a **Probability of False Negative in Detection of New Salting Techniques** under 10% along with an ability for **Fully Automated Creation of New Feature Extraction**. Section 2.0 describes these.

Also, in dealing with mimicry, given the growth and anticipated growth of white-listing, we explicitly state here that detection performance in detecting brand relevant messages is effectively equal to detection performance in detecting phishing. Any message pertaining to a phished brand is considered brand relevant if it seeks identity or account information either directly or indirectly through links..

Last, with the emphasis clearly on detection of such new and novel salting and mimicry techniques, we relax one constraint, namely “Labelled Content Processing Message Throughput.” In 2006, we set this goal to 12 messages per second. For 2007, to meet the ambitious new objectives described above for detection of new salting techniques, we relax this goal to 12 seconds per message. Relaxing this goal in this manner should facilitate focus on the core innovative and highly valuable fundamental research into adaptive and fully automated extraction of new complex features. The current prototype provides such salting detection with Labelled Content Processing Message Throughput of 5.4 seconds per message, so the newly relaxed goal provides ample margin for focus on practical but computationally intensive machine learning techniques. Even at the relaxed “per unit” performance goal, the parallelism of the Analysis System architecture could facilitate growing the Analysis System to full production capacity at a cost of roughly \$1M, well under the \$10M ceiling stated in 2006,

even if the hardware cost of a very large scale prototype could be \$100k, roughly twice the \$50,000 amount identified in last year's projections.

This summarizes the requirements stated last year and changes to those requirements made over the past year.

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**. Analysis System and Policy Enforcement Point are depicted and defined in greater detail in Section 3.0. 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

Regarding previously stated 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.

In dealing with mimicry, given the growth and anticipated growth of white-listing, please note that detection performance in detecting brand relevant messages is effectively equal to detection performance in detecting phishing. Any message pertaining to a phished brand is considered brand relevant if it seeks identity or account information either directly or indirectly through links.

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 by 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. NOTE: LATENCY depends on interactions between the centralized Analysis System and the distributed set of Policy Enforcement Points along with processing times at the Policy Enforcement Points and Analysis System respectively.

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.

New Performance Parameters:

Probability of False Positive in Detection of New Salting Techniques (**PFPNST**)
- The percentage of messages flagged by the Analysis System as having a previously unseen salting technique while lacking such a previously unseen salting technique divided by the number of messages lacking such a previously unseen salting technique sent through the Analysis System specifically to determine FFPNST after conclusion of a training period.

Probability of False Negative in Detection of New Salting Techniques (**PFNNST**)
- The percentage of messages with previously unseen salting techniques that are not detected as having previously unseen salting techniques divided by the number of messages with previously unseen salting techniques sent through the Analysis System specifically to determine PFNNST after conclusion of a training period.

Please note that FFPNST and PFNNST are being improved in the context of new salting techniques being exceptionally rare. Currently, new salting techniques appear in less than one message per million messages sent over the internet. However, for testing purposes, to have meaningful statistical confidence in metrics in such a challenging area, testing will be done on test sets where the percentage of messages containing previously unseen salting techniques is roughly the percentage of messages not containing previously unseen salting techniques.

Over the past year, the AntiPhish Consortium has developed the ability to detect individual messages with new salting techniques. This exceeds original hopes for being able to detect new salting techniques through clustering and performance analysis and other statistical analysis of sets of messages containing larger subsets of messages with new salting techniques. Being able to detect the first such message with new salting techniques could be exceptionally valuable, even if computationally intensive. Although such computationally intensive techniques may be well suited for the Analysis System, it would be ideal if the Analysis System could generate a heuristic for such feature extraction so that the heuristic could be run at each Policy Enforcement Point with less computational burden on the Policy Enforcement Points than running the techniques for detection of previously unseen salting techniques on every message at every Policy Enforcement Point.

For these reasons, and because such a heuristic could be far more precise than detection of previously unseen salting, we set a new higher goal of **Fully Automated Creation of New Feature Extraction (SYNTH)**. Because of the open ended nature of such early fundamental research, we express this as a general ability. Eventually, this may be measured in many dimensions. Such dimensions may eventually include the receiver operator characterization (ROC) of the synthesized feature extraction techniques, the number of sample messages required to achieve such ROC performance, and the computing burden required synthesize such feature extraction. However, setting specific goals in any of these areas would be premature at this point.

3 Architectural Context

The graphic below is provided for architectural context.

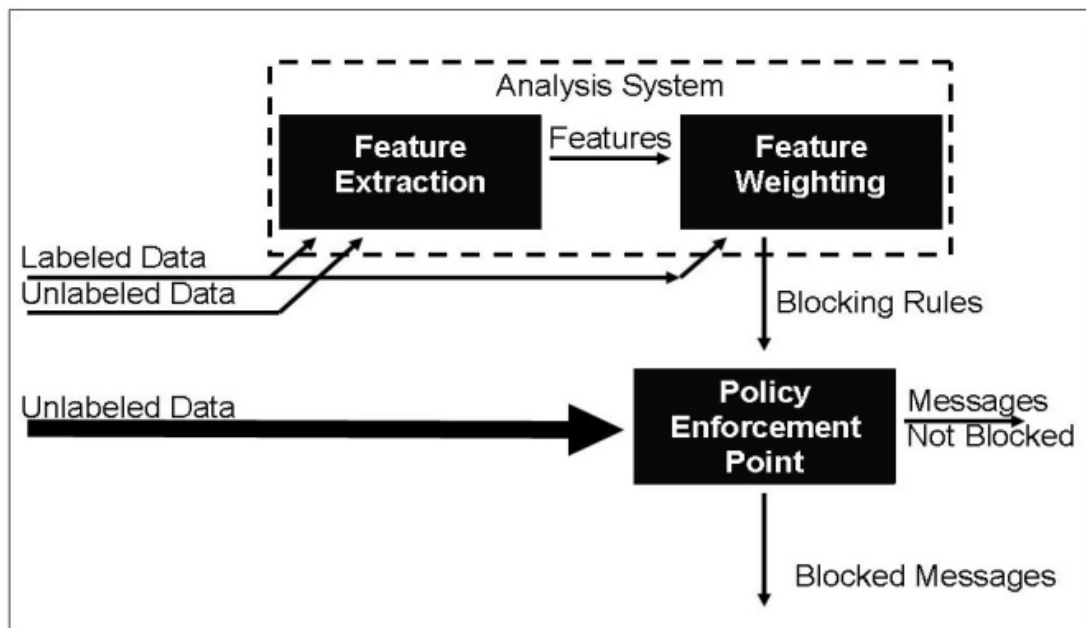


Figure 1: Architectural Depiction

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 Production	Prototype	Production	Goal
PFN	proprietary	0.10*	0.05	0.01
PFP	proprietary	0.01	0.001	1/1M*
RECALL	proprietary	0.90	0.95	0.99
PRECISION	proprietary	0.95	0.99	0.999999
LCPMT	proprietary	0.08/U	100	1,000
LCPVT	proprietary	0.12	12	100
LATENCY	300	3,000	300*	30
PERSONNEL	proprietary	1	20	5
HCAS	proprietary	TBD	2M	1M
HCPEP	proprietary	10k	1M	10k
PEPMT	proprietary	100,000	100,000	1M
PEPVT	1,000	1,000	1,000	10,000
PFNST	proprietary	0.01	0.001	0.00001
PFNST	proprietary	0.10	0.050	0.0001
SYNTH	No	Goal of Yes	Goal of Yes	Goal of Yes

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