



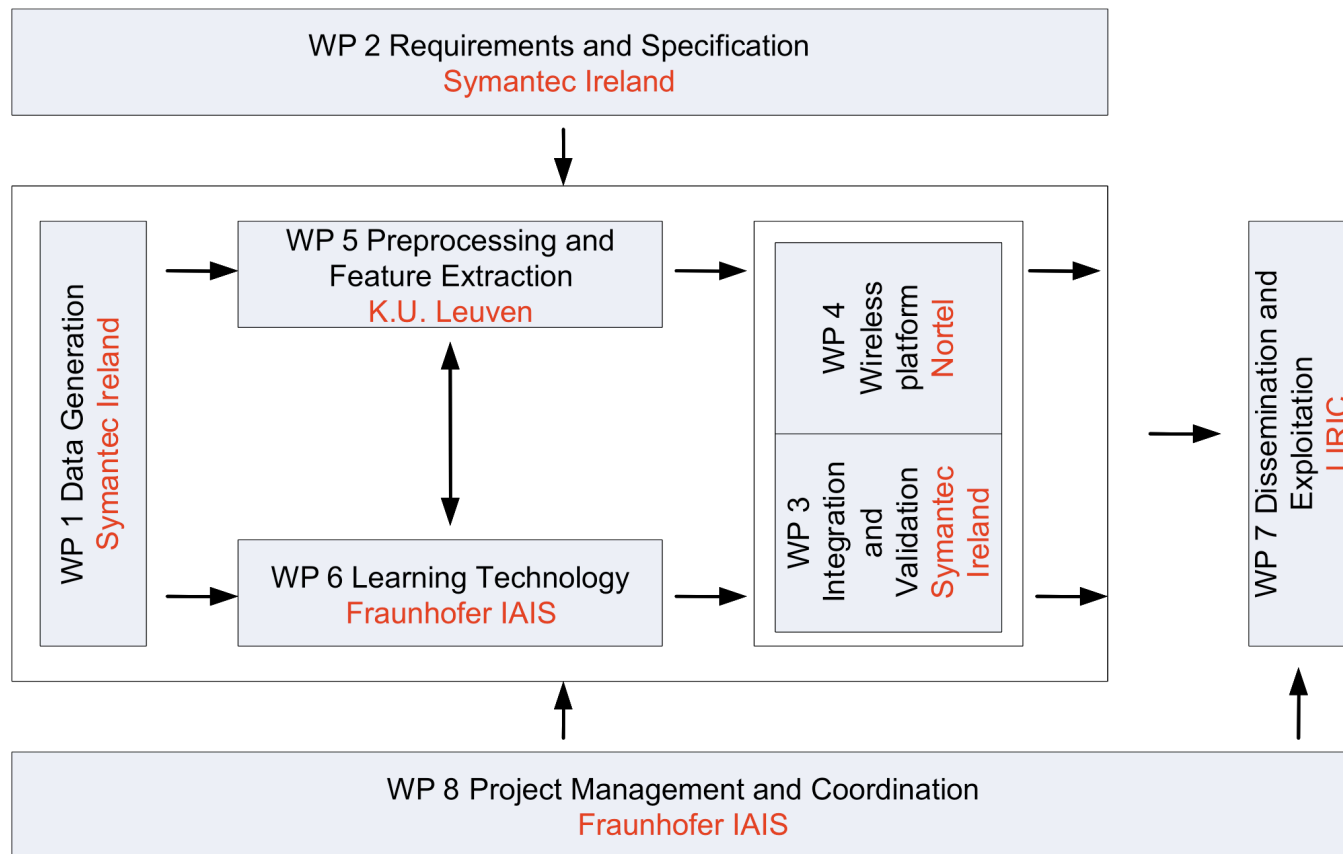
AntiPhish Project Presentation

Brian Witten

January 2008



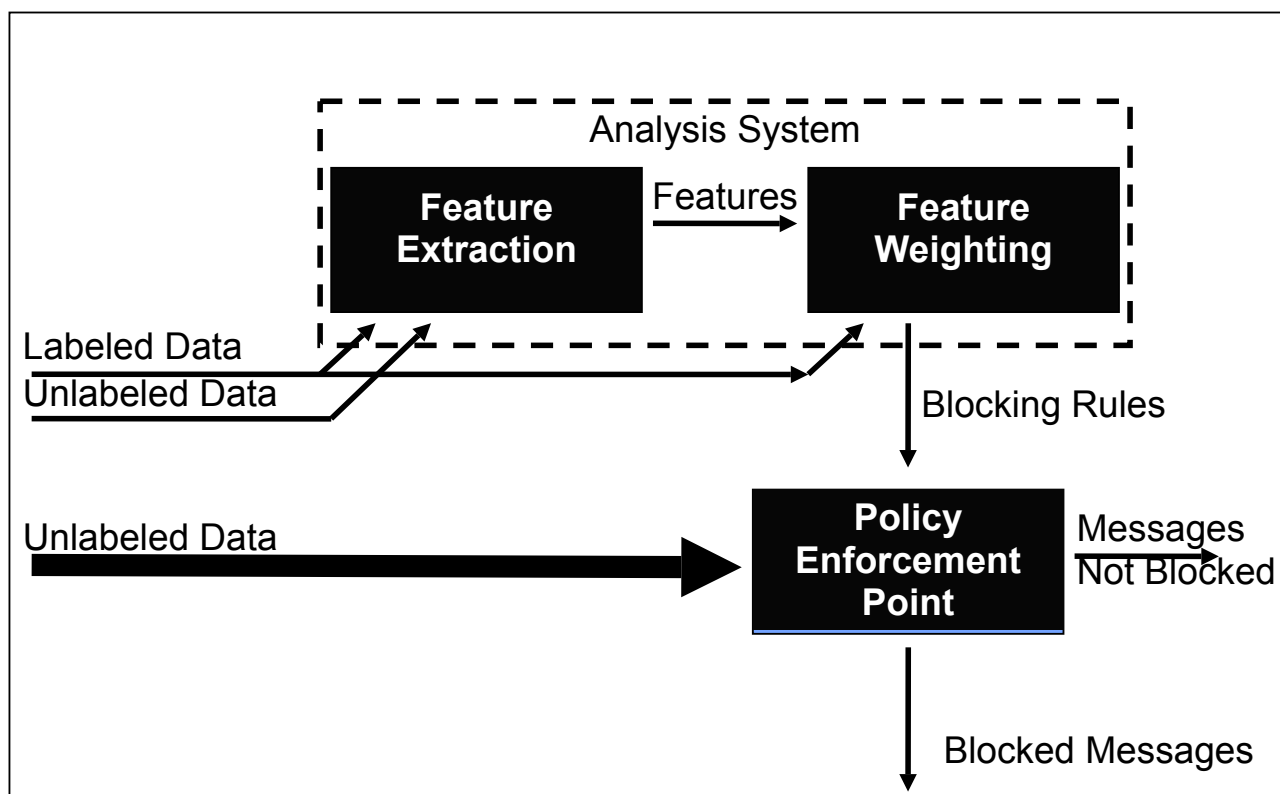
Work Package Structure



Agenda

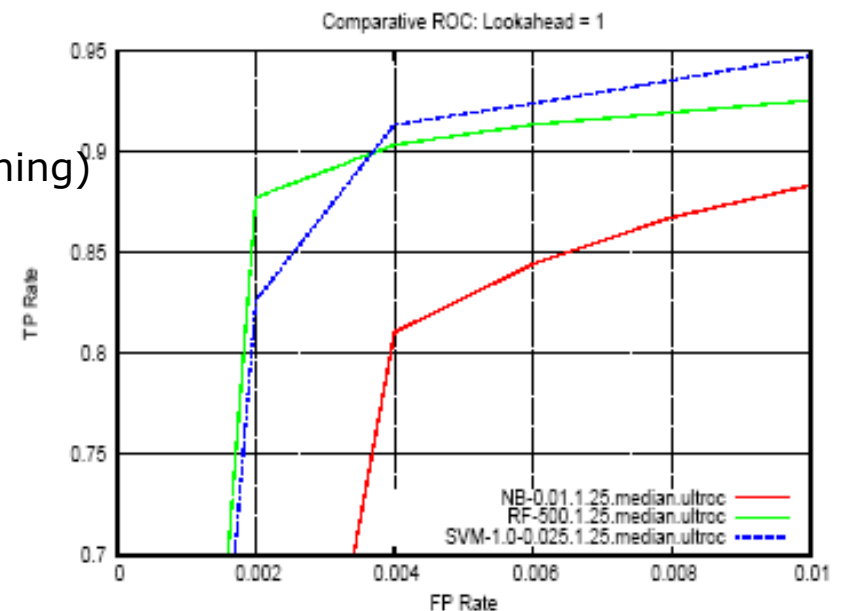
- 1. Work Package 2 – Requirements and Specification**
- 2. Work Package 1– Data Generation and Dissemination**
- 3. Work Package 5 – Message Pre-processing & Feature Extraction**
- 4. Work Package 6 – Advanced Learning Technology**
- 5. Work Package 3 – Integration and Validation**
- 6. Work Package 4– Wireless Platform**
- 7. Work Package 7– Dissemination and Exploitation**
- 8. Work Package 8– Project Management and Coordination**

Work Package 2 – Architecture Specification (Run Time Depiction)



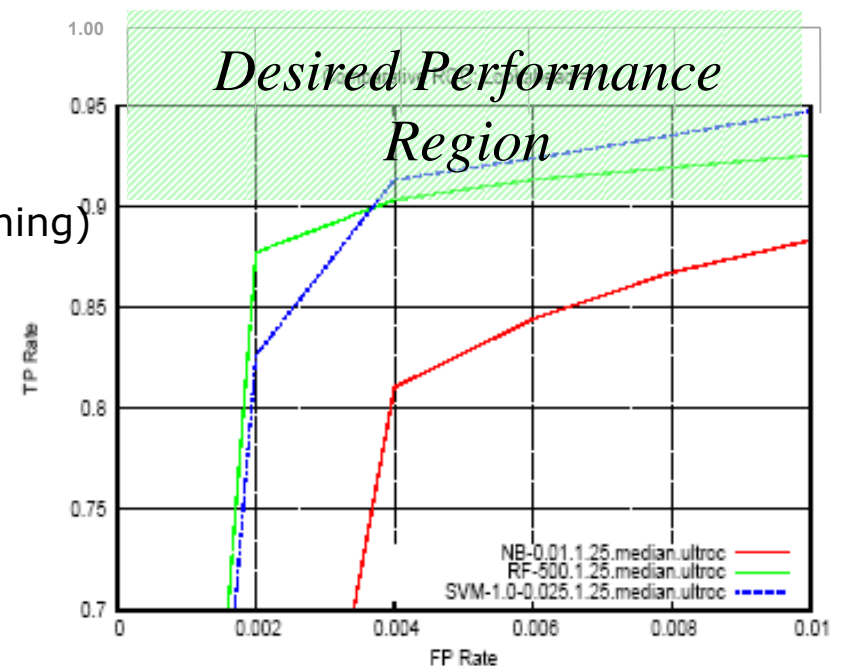
Work Package 2 – Performance Requirements

- Develop dynamic feature selection of sufficient quality to beat past performance of machine learning (ML) techniques, even where ML techniques were optimized with static feature selection.
- Performance Points:
 - A: Prototype (Phishing)
 - B: Production Requirement (Phishing)
 - C: Production Goal (Phishing)
 - D: Brightmail (Spam, Current)
- Additional requirements include number of messages per minute, volume per minute in megabytes, and with reasonable hardware and staff availability constraints.



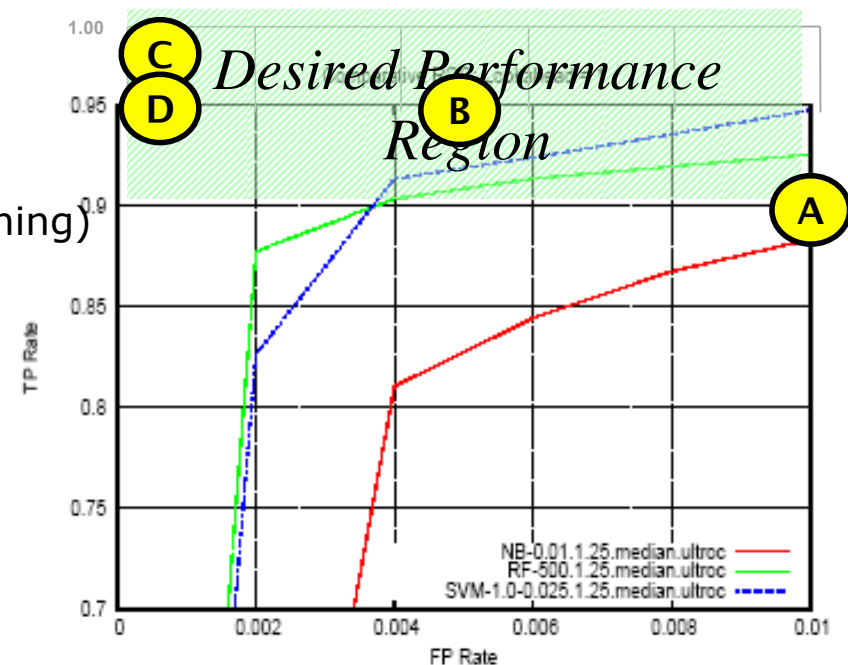
Work Package 2 – Performance Requirements

- Develop dynamic feature selection of sufficient quality to beat past performance of machine learning (ML) techniques, even where ML techniques were optimized with static feature selection.
- Performance Points:
 - A: Prototype (Phishing)
 - B: Production Requirement (Phishing)
 - C: Production Goal (Phishing)
 - D: Brightmail (Spam, Current)
- Additional requirements include number of messages per minute, volume per minute in megabytes, and with reasonable hardware and staff availability constraints.



Work Package 2 – Performance Requirements

- Develop dynamic feature selection of sufficient quality to beat past performance of machine learning (ML) techniques, even where ML techniques were optimized with static feature selection.
- Performance Points:
 - A: Prototype (Phishing)
 - B: Production Requirement (Phishing)
 - C: Production Goal (Phishing)
 - D: Brightmail (Spam, Current)
- Additional requirements include number of messages per minute, volume per minute in megabytes, and with reasonable hardware and staff availability constraints.



Work Package 2 - New Performance Goals

Fraunhofer and Leuven have now demonstrated the ability to detect previously unseen salting techniques. To begin quantifying progress of this important new capability, we have defined three new metrics and goals:

- Probability of False Positive in Detection of New Salting Techniques (PFPNST)
- Probability of False Negative in Detection of New Salting Techniques (PFNNST)
- Fully Automated Creation of New Feature Extraction (SYNTH).

	Current Production	Prototype	Production	Goal
PFPNST	proprietary	0.01	0.001	0.00001
PFNNST	proprietary	0.10	0.050	0.0001
SYNTH	No	Goal of Yes	Goal of Yes	Goal of Yes

Work Package 1– Data Generation

- First Year:
 - Privacy Agreements completed
 - First dataset delivered
- Second Year, additional datasets provided:
 - Third training dataset: roughly 22.5 million emails
 - 173GB delivered to Fraunhofer January 2008
 - Data spans December 29 2006 to November 2007

label category	Message count	%
spam (regular)	21,455,404	96.427
spam (brand relevant)	442,677	1.989
spam (confirmed phishing/fraud)	328,440	1.476
ham (newsletters)	23,954	0.107

- Ratio of Phishing to other spam is consistent with ratios found in wild, but proportion of Ham is artificially low due to privacy concerns.

Work Package 5: Message Preprocessing / Feature Extraction

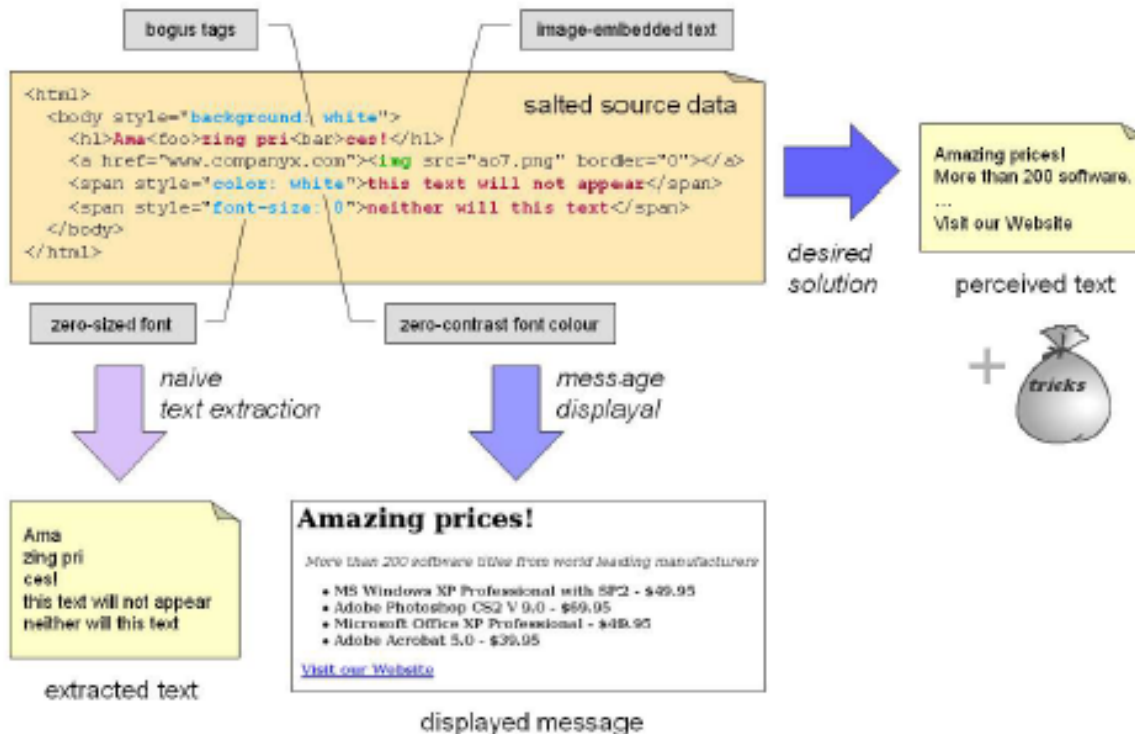


Illustration 1: The problem of text salting (left), and the desired solution (right)

Work Package 5: Message Preprocessing / Feature Extraction

- KU Leuven developed techniques for detecting previously unseen evasion techniques. These techniques have been submitted for publication, "Transactions On Information Systems (TOIS) ACM Journal [De Beer and Moens, 2008]"
 - *These techniques have also been submitted to patent authorities in application for a patent.*
- Syntactic Features now include orthographic, typographic, morphological, character usage, character n-grams, lexical structure (paragraphs, sentences, words), part of speech tags, stemmed words, aggregate word index.

Work Package 5: Message Preprocessing / (Semantic) Feature

- Fraunhofer has now implemented:
 - PLSA (Probabilistic Latent Semantic Analysis)
 - LDA (Latent Dirichlet Allocation)
 - NMF (Non-negative Matrix Factorization)
 - CLTOM (Class Topic Model)
 - Word unigram model.
 - Last, a specific version of the Dynamic Markov Chain (DMC) treats an email as a sequential stream and approximates a probabilistic generator for this stream. Phishing, spam and ham generators were trained. The resulting likelihood scores were used as an additional input to general classifiers such as SVM to boost the performance of such classifiers.

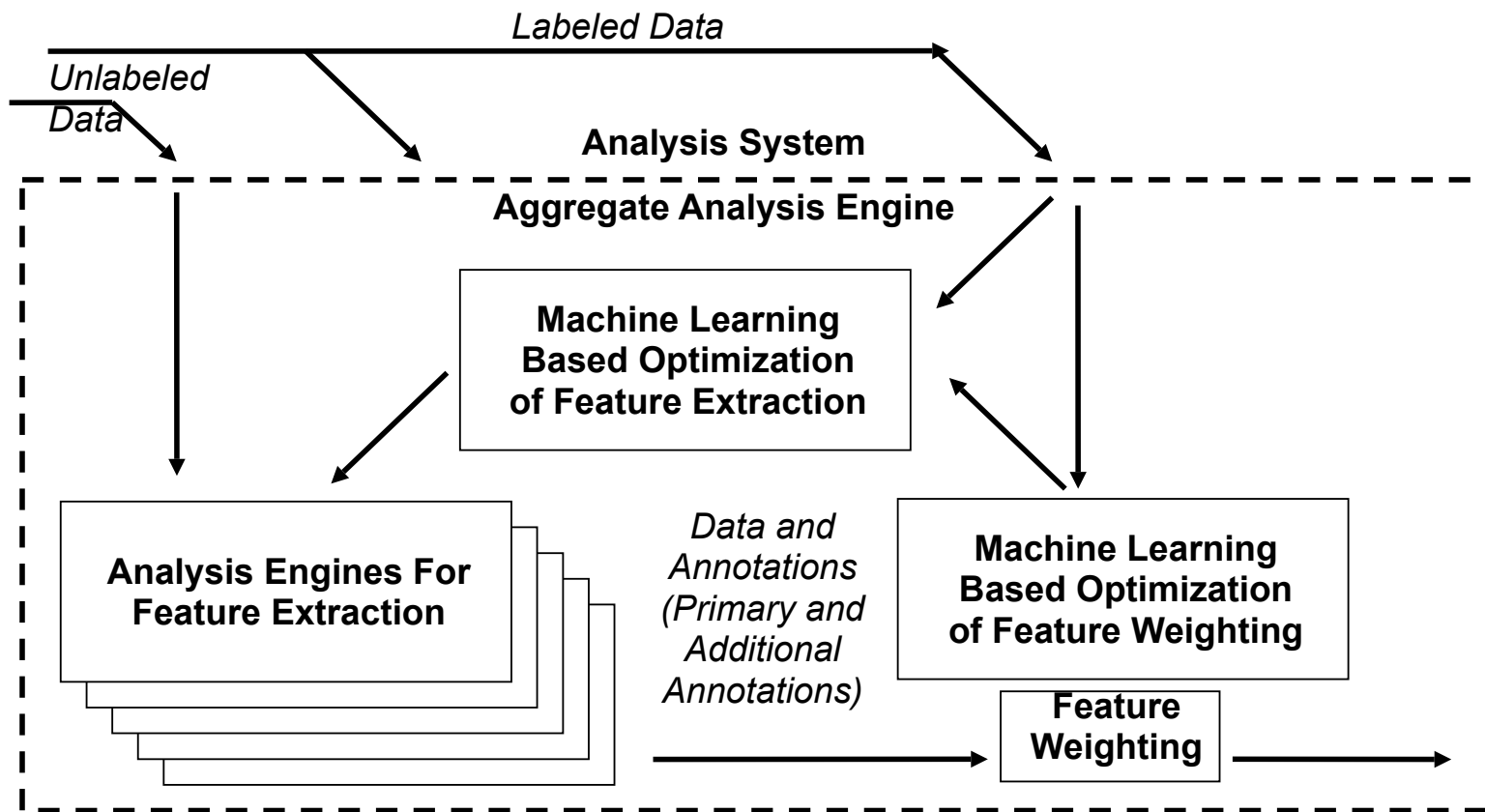
Work Package 5: Message Preprocessing / Feature Extraction

- Structure and layout feature extraction has been completed by Fraunhofer and KU Leuven in collaboration.
- Given the shifts to image spam mentioned in the last review, fast and efficient handling of images is crucial.
 - Fraunhofer integrated a fast C++ Imaging Library which they had developed separately.
 - After evaluation of GNU Optical Character Recognition (GOOCR), OCRad, and ABBYY FineReader, Fraunhofer integrated ABBYY FineReader which performed best.
 - Symantec provided an implementation of logo detection.

Work Package 6 – Advanced Learning Technology

- Fraunhofer completed development and integration of a prototype which has now been validated in testing by Symantec
- Review of Algorithms and Workflows
 - On-line learning with kernels (Kivinen et al 2001)
Very efficient for very high dimensional learning.
 - Perceptron, MIRA [Crammer 04,06], LASVM [Bordes et al. 05],
 - L2-SVM [Keerthi, DeCoste 05]
square loss, 400-fold speed increase vs. usual SVM
 - SVMPerf [Joachims KDD06], Transductive SVM [Joachims 1999]
 - Multinomial model [Nigam et al 1999], self training [Yarowski 1995]
- Innovation:
 - Cascading Classifiers perform at SVM levels but with computational effort reduced by 50% to 70%
 - 92% f-value detecting previously unseen evasion techniques
 - 29% better Phish detection (f-value increased from 94.4% to 96%)

Work Package 3 –Integrated Architecture



Work Package 3 – Validation

1. Phishing Classification

Performance Points:

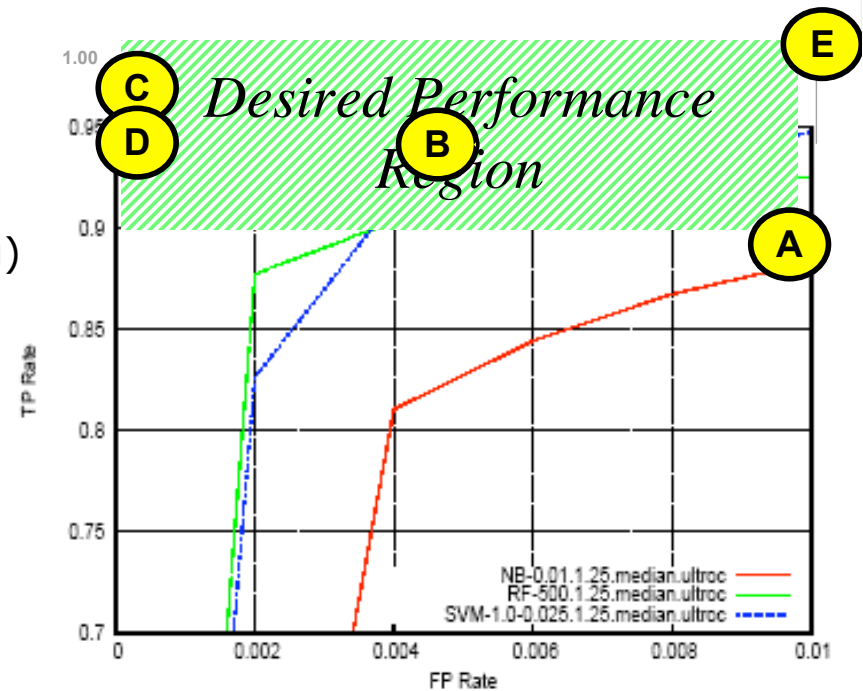
- A: Prototype (Phishing)
- B: Production Requirement (Phishing)
- C: Production Goal (Phishing)
- D: Brightmail (Spam, Current)
- E: Initial Prototype Tests

	classified as phishing	classified as non-phishing	actual totals
phishing	2387	70	2457
Non-Phishing	72	2352	2424
Totals	2459	2422	4881

Table 2. Trial results - Confusion Matrix

PRECISION	0.970
RECALL	0.971
PFP	0.029
PFN	0.028

Table 3. Trial results - Metrics



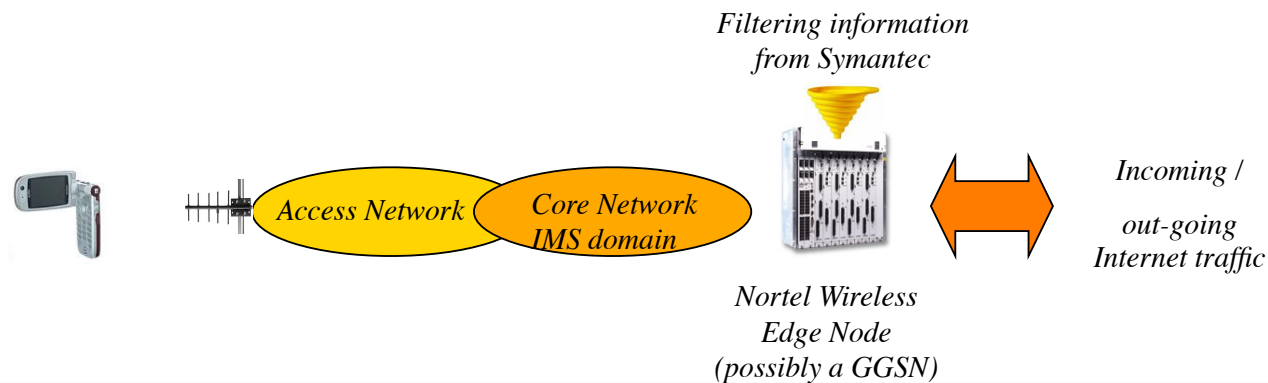
Work Package 3 – Validation (continued)

2. New Evasion Detection

- Confirmed: Prototype detects previously unseen evasion techniques
 - Methodology:
 - Remove messages with a given salting technique from the training corpus;
 - Test ability to detect previously unseen salting techniques
 - Correctly flagged all tested salting techniques, including:
 - Font colour, font size, read order, clipping and concealment.
 - Due to limitations of OCR:
 - Not possible to precisely access ROC performance of new evasion detection.
 - This was strictly prior to ABBYY integration.
 - Testing with ABBYY integration in progress now.

Work Package 4– Wireless Platform

- Apply AntiPhish techniques to Wireless environment, with a specific focus on legacy 3GPP network architectures
- Demonstrate applicability to the ever growing wireless traffic, including mail, SMS, MMS, ...
- Current architecture is:
 - access type agnostic (e.g. 2G/GSM, 3G/UTRAN, 4G/LTE and possibly WLAN access)
 - compatible with 3GPP upcoming "Enhanced Packet Core" evolutions being studied



Work Package 7 – Dissemination and Exploitation

- On December 4, 2006, Symantec issued a press release on behalf of the consortia members with their approvals
- Licensing agreements are established amongst the partners for the commercial exploitation of the research
- Patent applications have been filed to protect the intellectual property
- Papers submitted for publication include:
 - “Detection of previously unseen evasion techniques,” Transactions On Information Systems (TOIS), ACM Journal [De Beer and Moens, 2008]
 - “New Features for Phishing Email Detection,” Submitted to SigIR 2008 [Bergholz, Chang, Paaß, Reichartz, Strobel 2008]

Work Package 8 – Project Management and Coordination

- Schedule of Meetings Held in 2007
 - Leuven, BE (10.01.2007)
 - St. Augustin, DE (16.02.2007)
 - Dublin, IRL (23.-24.05.2007)
 - St.Augustin, DE (22.-23.08.2007)
 - Dublin, IRL (15.-16.10.2007)
- Bilateral Meetings
- Coordination is also done through e-mail mailing lists, a private web server providing Basic Support for Collaborative Work (BSCW), and monthly teleconferences.
- Changes in Participation
 - Responsibility for the Nortel participation shifted from Pierre Lescuyer to Thomas Genouville and later to Bob Smith

Summary

- Concluding the second year of a three year effort
- Symantec has validated key capabilities of the prototype developed by Fraunhofer and K.U.Leuven
 - Detection of previously unseen evasion techniques
 - ROC performance near other world leading systems, but without manual selection of narrow feature set
- Spam and Phishing threats continue adapting quickly, but the original goals remain crucial, and the consortia still adapts emphases in tracking long term changes
- Emphases for coming year will be field tests with Tiscali, wireless testing with Nortel, and disseminating results