

**SYMBOLIC**

NETWORK SECURITY COMPANY™

# Strategie per affrontare lo Spam

Fabrizio Cassoni  
fc@symbolic.it

## C'è Spam e SPAM

- Il termine proviene da un prodotto Fast Food
- Il termine “spam” associato a qualcosa di fastidioso, che impedisce una normale conversazione, deriva da uno sketch dei Monty Python
- Una definizione piuttosto calzante dello Spam è “unethical mass-mail”



## Mezzi di trasmissione

- Lo Spam raggiunge gli utenti tramite:
  - Usenet
  - E-mail individuale
  - Mailing list
- Il futuro:
  - SMS?
  - Pop-Up (si tratta di Spam?)

## Magnitudo del problema

- AOL dichiara di bloccare ca. 67 junk mail al giorno per ognuno dei propri utenti  
([http://biz.yahoo.com/bw/030430/305356\\_1.html](http://biz.yahoo.com/bw/030430/305356_1.html))
- Microsoft, AOL e Yahoo sono state costrette ad affrontare il problema dello spamming effettuato tramite i rispettivi servizi di e-mail
- Il centro anti-frode del FBI ha ricevuto 48.000 denunce di frodi perpetrate da spammers nel 2002
- Secondo alcune fonti, si potrebbe arrivare a una percentuale di spam del 50% su tutto il traffico di e-mail

## Perché è un problema

- Tempo perso per scaricare/leggere/cancellare i messaggi
- Rischio di truffe per chi risponde
- Perdita di privacy per chi chiede di essere cancellato dalla “mailing list”
- Rischi legali per le aziende, se i propri impiegati fanno il forward dello spam
- Eccessivo carico di lavoro per i server
- Traffico di rete inutile

## Rischi a lungo termine

- L'eccessiva quantità di spam nella posta degli utenti può portare a:
  - diminuzione dell'importanza della posta elettronica come strumento di lavoro
  - diffidenza e disaffezione verso l'e-mail
    - La posta elettronica è stata la ragione primaria del successo di Internet presso tutte le categorie di utenti

# Non c'è un solo tipo di Spam

- Proposte di carattere commerciale
  - Multi-level marketing, investimenti, offerte di prodotti
- Pornografia
- Hate Mail
- Chain letters
- Friendly Spam
  - Barzellette, animazioni, scherzi inviati da conoscenti
- Hoax
  - allarmi riguardanti virus inesistenti
  - leggende metropolitane

# Le soluzioni secondo SurfControl

- SurfControl E-Mail Filter permette di combattere lo Spam in diversi modi:
  - Anti-Spam Agent
  - Spam Treshold
  - LexiMatch
  - Virtual Learning Agent
  - Virtual Image Agent
  - HTML Stripper
  - Real-Time Black List
  - Deny list
  - Reverse Client DNS Lookup
  - Closed relay

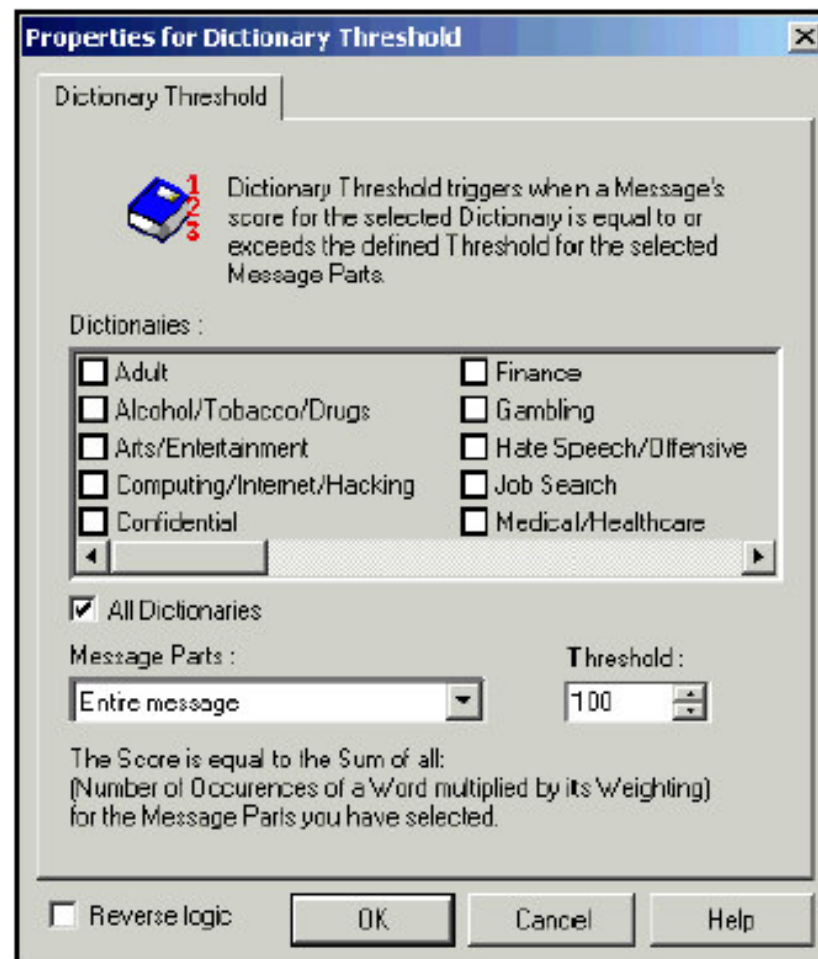
# Anti-Spam Agent

- Database di “firme” di junk mail note
- Lo Spam è suddiviso in diverse categorie (Materiale per Adulti, Hoax, Chain Letter ecc.) in modo da consentire una gestione granulare
- I database vengono costantemente aggiornati
- Vengono esaminati anche gli allegati, non solo il testo della mail
- 100% di accuratezza sullo Spam noto



# Spam Threshold

- Basato su dizionari di termini comunemente trovati nello Spam
- Multilingue
- E' possibile assegnare un "peso" a ogni termine
- Ogni ricorrenza delle parole classificate nella mail aumenta il "punteggio" totale
- Al raggiungimento di una soglia prefissata, la mail viene considerata Spam




# LexiMatch

- Utilizza i Content Dictionaries
- Consente di costruire filtri che applicano operatori booleani (AND, OR, NOT, NEAR)
- Permette l'uso di wildcard e pattern alfanumerici
- E' in grado di intercettare variazioni di grafia sui termini ricorrenti nello Spam

**Properties for LexiMatch**

LexiMatch

Near Distance:  LexiMatch triggers when the LexiMatch Condition evaluates to TRUE. Click the Condition Builder button below to match words, phrases or binary pattern. The Near Distance is the number of characters between two words.

80

Message Parts: Whole Message

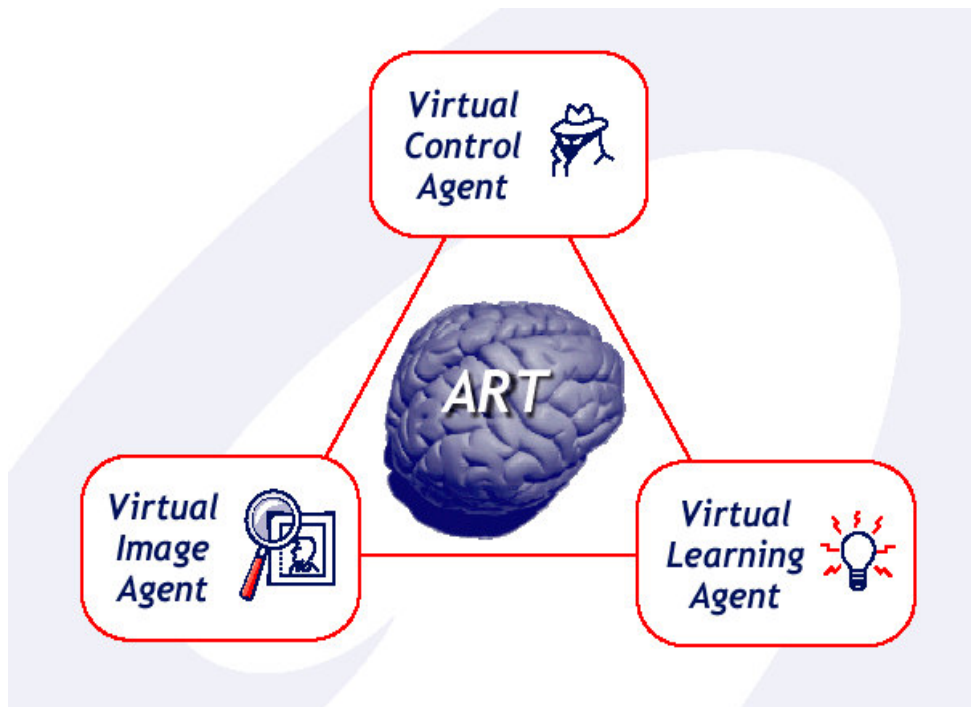
Dictionary: Adult

Join	Word 1	Operator	Word 2
SCAN FOR	breast	OR	bosum
OR NOT	breast	NEAR	cancer
OR NOT	breast	NEAR	chicken
OR NOT	sexual	NEAR	education

Reverse logic

OK Cancel

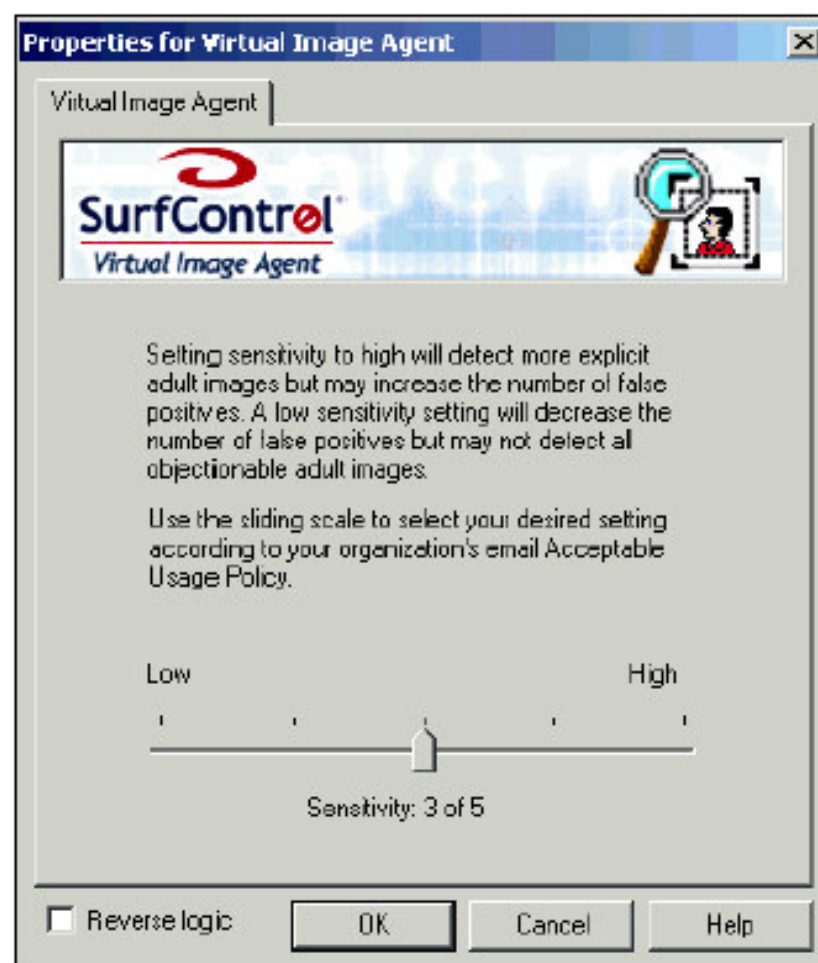
# Adaptive Reasoning Technology



- Tecnologia basata su Reti Neurali
- Implementata in Web Filter e E-Mail Filter
- In grado di permettere ai prodotti di “imparare”
- Permette di categorizzare automaticamente il nuovo contenuto

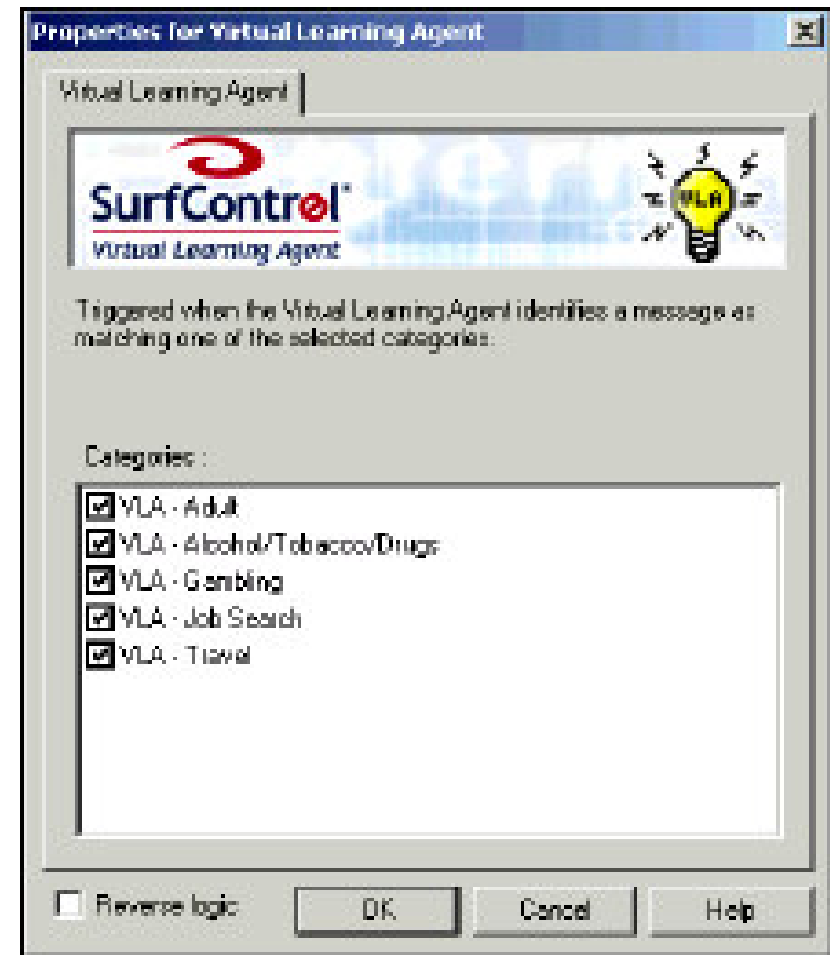
# ART - Virtual Image Agent

- Strumento basato sulla Adaptive Reasoning Technology
- In grado di esaminare le immagini allegate alle mail
- Basandosi su 22.000 algoritmi di Intelligenza Artificiale, è in grado di riconoscere materiale grafico “per adulti”
- Sensibilità regolabile
- Efficace per bloccare Spam pornografico e Friendly Spam di cattivo gusto

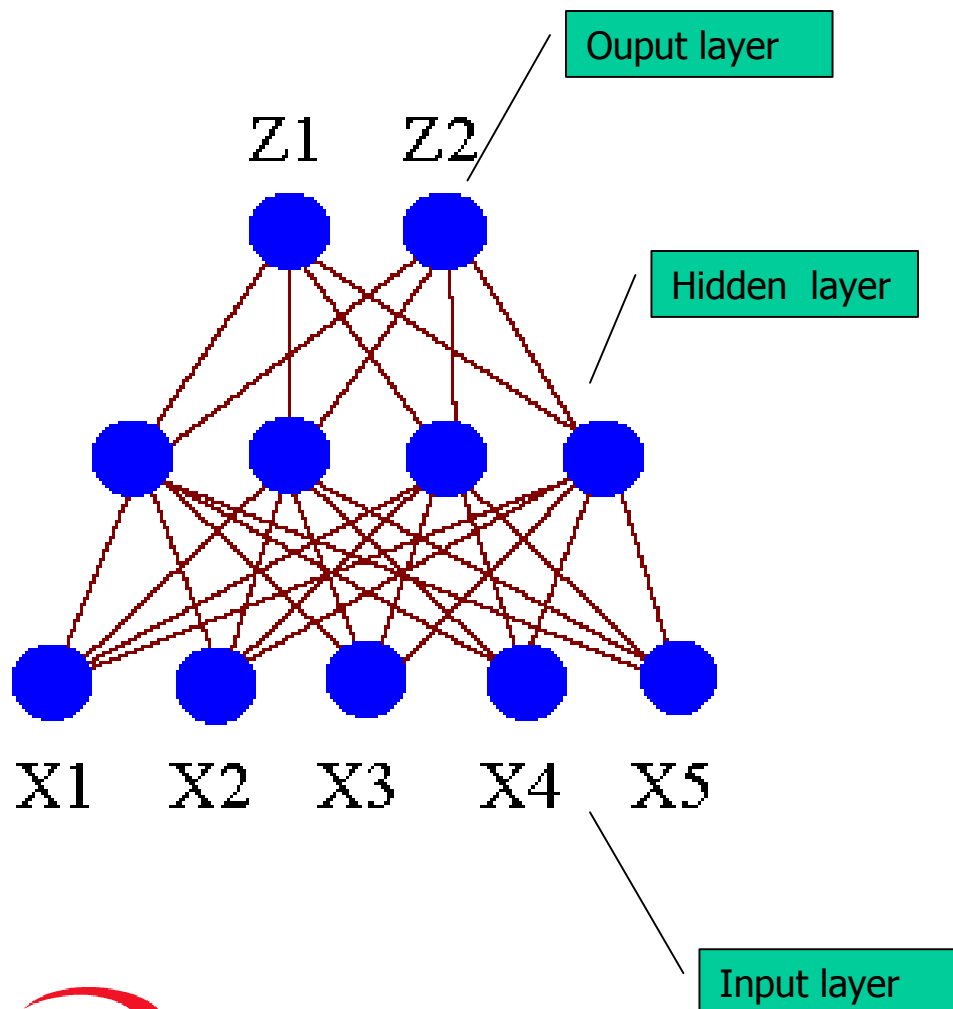


# ART - Virtual Learning Agent

- VLA applica il concetto di Reti Neurali per categorizzare il contenuto
- Richiede un “training” iniziale per associare frasi, parole e pattern a una categoria
- Il compito di VLA è di analizzare parole e frasi nel contesto appropriato



# Principi di base di VLA



- Una Rete Neurale è costruita su una serie di nodi che cercano di simulare il comportamento dei neuroni
- Un input viene fornito alla Rete Neurale, assieme a un set di dati storicizzati con cui confrontare i risultati dell'elaborazione
- Nello hidden layer, la Rete effettua una somma ponderata dei dati ricevuti nell'input layer
- Lo scopo è di avere un meccanismo che "impara" continuamente dall'esperienza, cioè che confronta i dati elaborati nell'hidden layer con i dati storici
- La qualità del risultato dipende largamente dal materiale di training

## Il training di VLA

- Al VLA vengono sottoposti messaggi di esempio, appartenenti a categorie predeterminate
- Vengono forniti sia l'input che l'output layer
- Per ogni esempio, utilizzando gli algoritmi adattativi su cui è costruito, VLA cerca di categorizzare la mail e confronta il suo risultato con quello prestabilito
- Se il risultato non corrisponde a quello predefinito, VLA aggiusta i suoi parametri fino ad ottenere il matching
- Da questo punto in poi, VLA applica quello che ha appreso ai nuovi messaggi, cercando di inserirli nella categoria corretta

## Conclusioni

- Il problema dello Spam è in forte crescita
- Non esiste ad oggi una singola tecnologia in grado di fornire un deterrente
- Un approccio multi-layered è la migliore strategia
- SurfControl fornisce soluzioni basate su dizionari, analisi lessicali e meccanismi di auto-apprendimenti per offrire agli amministratori un insieme di strumenti flessibili e intelligenti
- Gli strumenti SurfControl sono in grado di adattarsi alle nuove tattiche degli Spammer e consentire agli amministratori di rete una risposta adeguata al problema