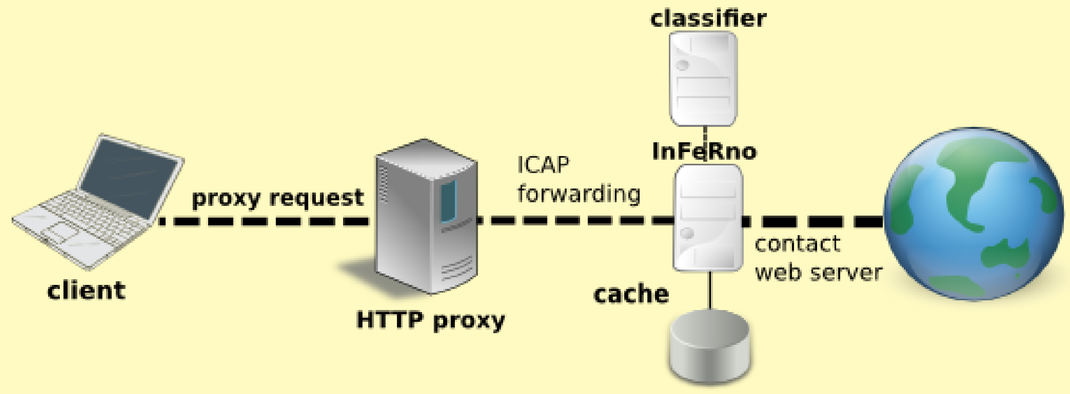
InFeRno - an Intelligent Framework for Recognizing Pornographic Web Pages

Sotiris Karavarsamis, Nikos Ntarmos and Konstantinos Blekas
University of Ioannina, Greece
{cs061205, ntarmos, kblekas}@cs.uoi.gr

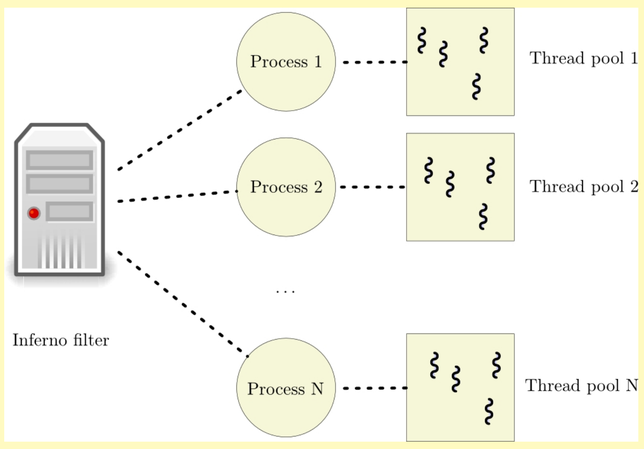
Problem preliminaries

- We propose *InFeRno*, an intelligent web pornography elimination system, classifying web pages based solely on their visual content
- Despite the plethora of useful information scattered on the Web, the Internet has become a hostile place for unprotected people like children
- The main characteristics of our system are
  - A minimal but powerful vector space
  - An extra 'bikini' class that is observed to improve SVM performance
  - A highly accurate and fast classification system
  - A full-fledged ICAP-based (Internet Content Adaptation Protocol) implementation of the pornography elimination system

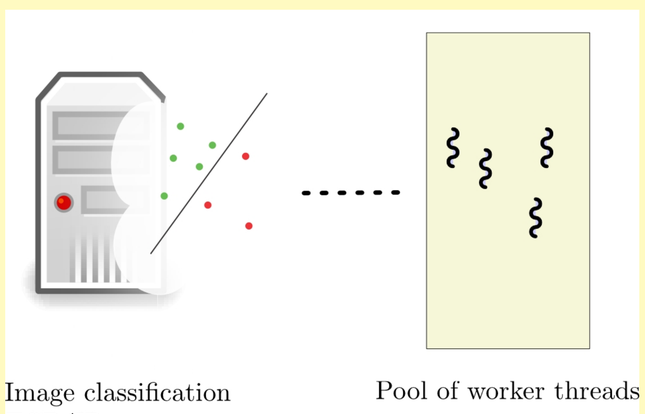
InFeRno architecture



- Standalone implementation of the InFeRno core as an ICAP module (integrates well with most HTTP proxy servers)
- Decoupled image classification and web page pre-processing (network I/O, web content fetching and web page fusion)
- A fast ISAM-based cache is implemented for fast cache I/O (classification lookups, updates, etc.)
- The administrator can tweak classification and/or network parameters, thus allowing for a flexible setting



InFeRno process/thread pool

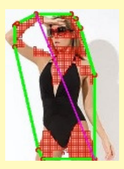


classifier thread pool

Classification system

original image

skin detection



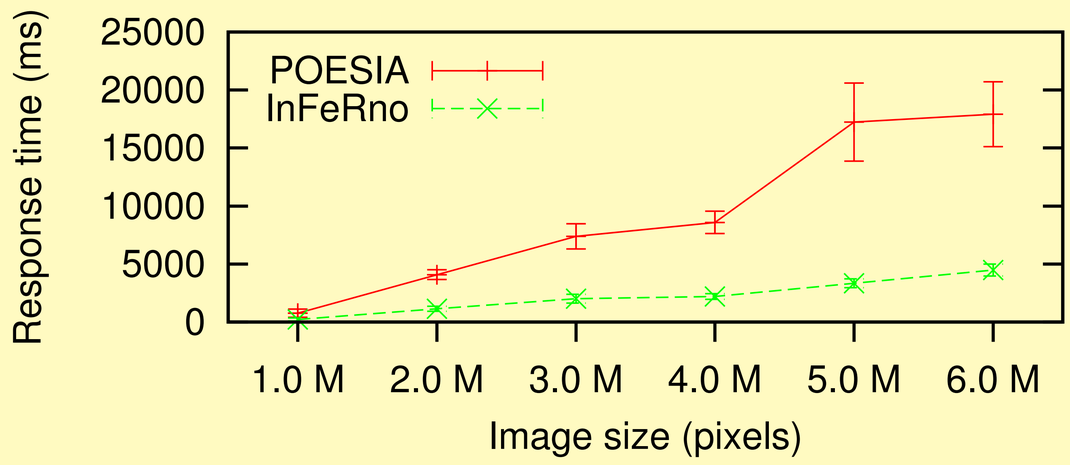
contour extraction

Three stages

- Skin detection [3] (Rule-based)
- Contour extraction [2] (region splitting scheme)
- Feature extraction and Classification
  - Extracted 15 features: RGB color statistics, skin-to-nonskin ratio, contour orientation, Hu moments
  - SVM classifier with RBF kernel

Experimental results

- Training dataset: manually collected 680 bikini images, 660 porn images and 4260 benign images from the Web
- Comparison with the *POESIA filter package* [1]
- Results:



  - $4\times$ **speedup** improvement
  - Accuracy: 98% for porn, 97% for bikini, 98.8% for benign class (comparable to POESIA filter)

References

1. POESIA filter project, "Poesia project" http://www.poesia-filter.org/.
2. W. Hu, O. Wu, Z. Chen, Z. Fu, and S. Maybank, "Recognition of pornographic web pages by classifying texts and images", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pp. 1019-1034, 2007.
3. V. Vezhnevets, V. Sazonov, and A. Andreva, "A survey on pixel-based skin color detection techniques," in *Graphicon*, pp. 85-92, 2003.