

Analyzing LISYS

Fernando Esponda, Stephanie Forrest and
Paul Helman

Dept. of Computer Science

Univ. of New Mexico.

fesponda@cs.unm.edu



Introduction

- Artificial Immune Systems for anomaly detection:
 - pH (process Homeostasis) (System Call Traces).
 - LISYS (TCP Connections).
- Ideas from vertebrate immune systems:
 - Protect from novel diseases by detecting the anomalous.
 - Learning without counter-examples.
 - Distributed, no centralized control.
 - Robust (no single point of failure) (population diversity).
 - Adaptive (environment dependent).



Components

- Problem:
 - Collection of fixed length binary strings we wish to protect (self).
 - Only a sample is available.
- Partial Matching (detectors):
 - r -contiguous bits.
 - r -chunks.
 - Computational considerations.
- Positive and Negative detection:
 - Tradeoffs: detector sets.
- Generalization:
 - Which and how many patterns are considered normal?
- Representation and Diversity:
 - What is a suitable representation?
- Response and Dynamic behavior.



Partial Matching Rules

- r -contiguous bits, $r=3$, $l=5$, windows=3:

- String 00111
- Detector 10110

- r -chunks

- String 00111
- Detectors 011

- Decomposition:

r -cb detector: 10110

r -chunk detector 1: 101

r -chunk detector 2: 011

r -chunk detector 3: 110

- Easier to analyze
- Implementation advantages
- r -chunks subsumes r -contiguous bits

Expected Number of Detectors

- The expected number of distinct patterns in a window of size r for a self set S :

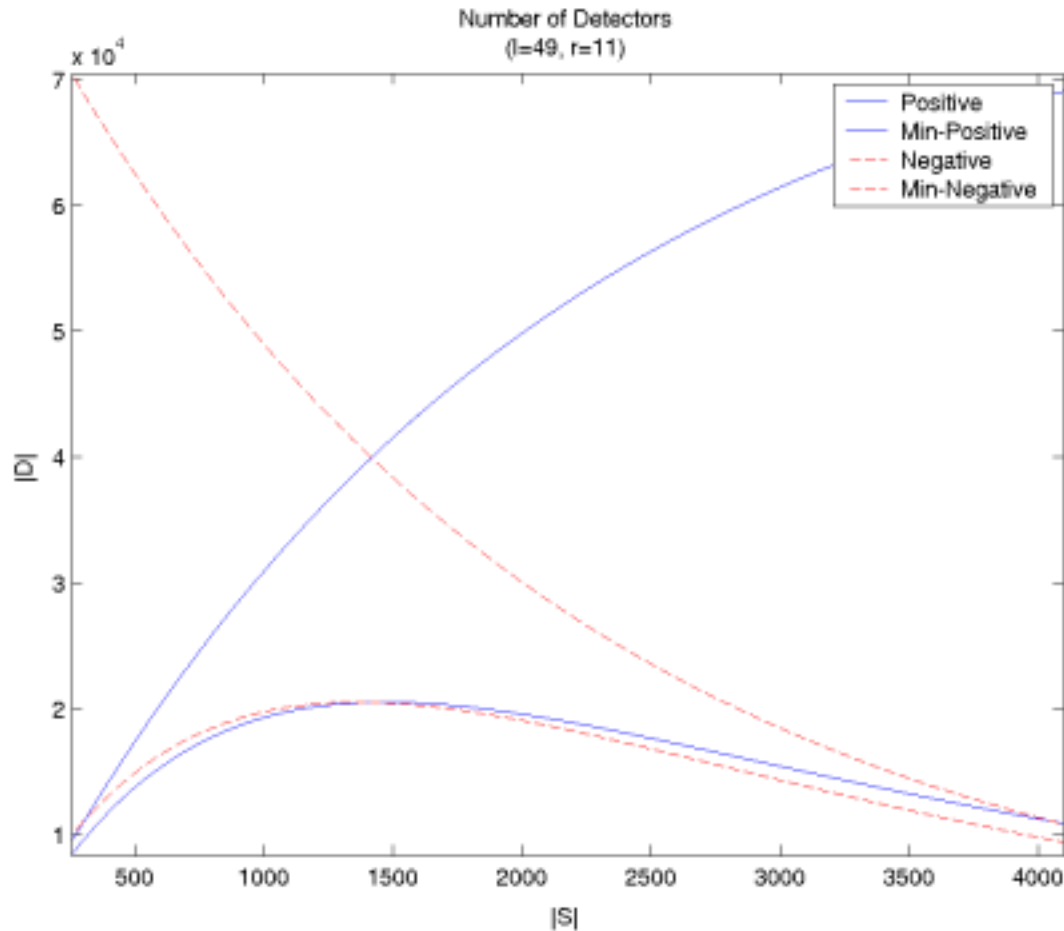
$$E_r = 2^r - 2^r (1 - 2^{-r})^{|S|}$$

- Expected number of detectors:
 - Positive: $E_p = t (E_r)$ where $t = l - r + 1$
 - Negative: $E_n = t (2^r - E_r)$

- $E_p = E_n = \frac{-1}{\log_2(1 - 2^{-r})} \in [2^{r-1}, 2^r]$



Detector Set Size



- The intersection is independent of the length of the strings.
- Minimal sets

Minimum set of detectors

- Redundancy:
 - Everything a detector matches is matched by some other detector.

$$E_{N_{\min}} = 2^r - E_r + (l - r)(E_r - 2(E_r - E_{r-1}))$$

- Implicit Matches:
 - Everything a detector matches is implied by other detectors.

$$E_{P_{\min}} = E_r + (l - r)(E_r - 2(E_r - E_{r-1}))$$



Generalization

- Window crossover
 - Two consecutive windows crossover if they match in their $r-1$ common bit positions. Ex. $r=2$:

$s1:$ 0000

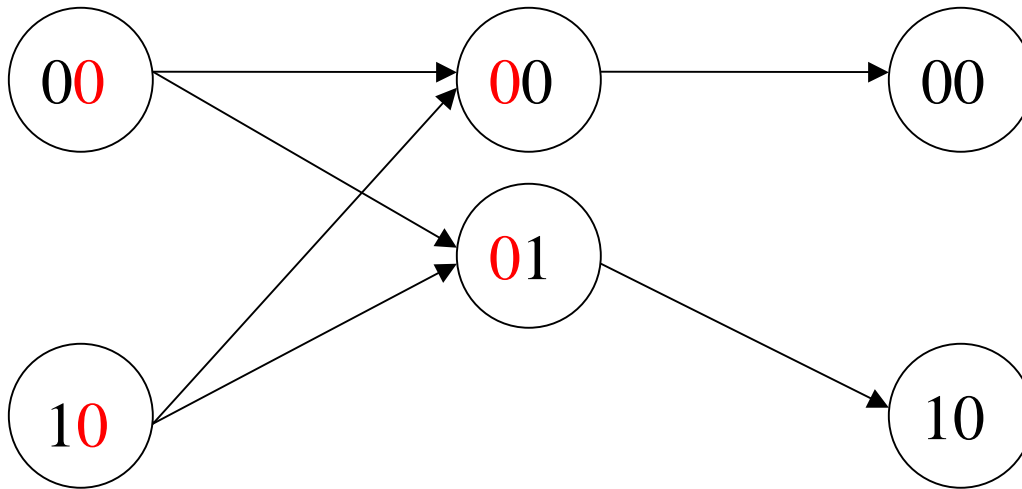
$s2:$ 1010

yield: 0010 and 1000

- The set of all crossover strings is the crossover closure CC .
- The set of protected strings by these schemes is exactly $CC(S)$.

Crossover Closure

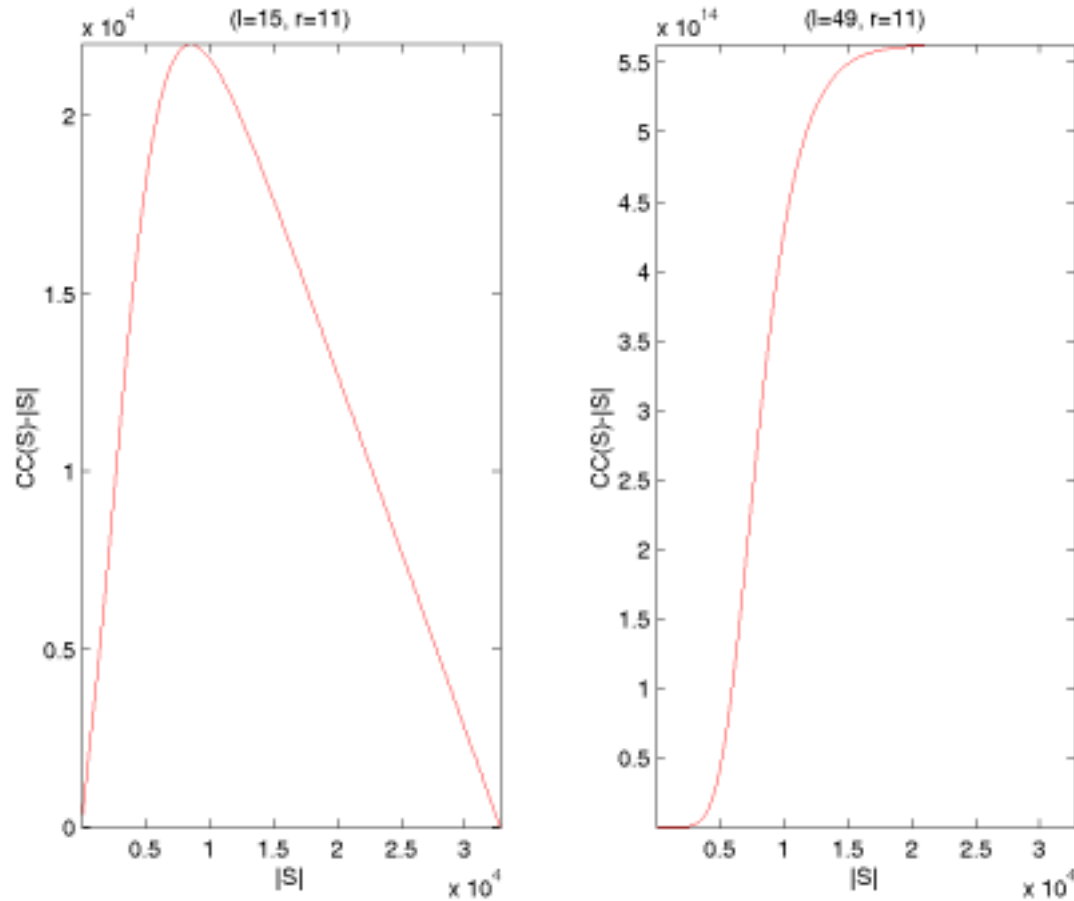
- $S = \{0000, 1010\}$, $CC(S) = \{0000, 0010, 1000, 1010\}$



$$CC(t) = \begin{cases} E_r, & \text{if } t = 1 \\ 2CC(n-1)P(2) + CC(n-1)(1 - P(2)), & \text{otherwise} \end{cases}$$

$$CC(t) = E_r (1 + P(2))^{t-1}$$

Crossover Closure



- The smaller r is in relation to l the quicker $CC(S)$ grows.
- Restrict sample size

Conclusions

- Step towards understanding and improving the systems we have already built:
 - Match rule
 - Positive and Negative detection
 - Detectors
 - Generalization
- Other work:
 - Representation (permutation masks)
 - Distance from $CC(S)$
 - Dynamic Behavior
- Biological metaphors:
 - Inspiration
 - Understand/Formalize
 - Depart

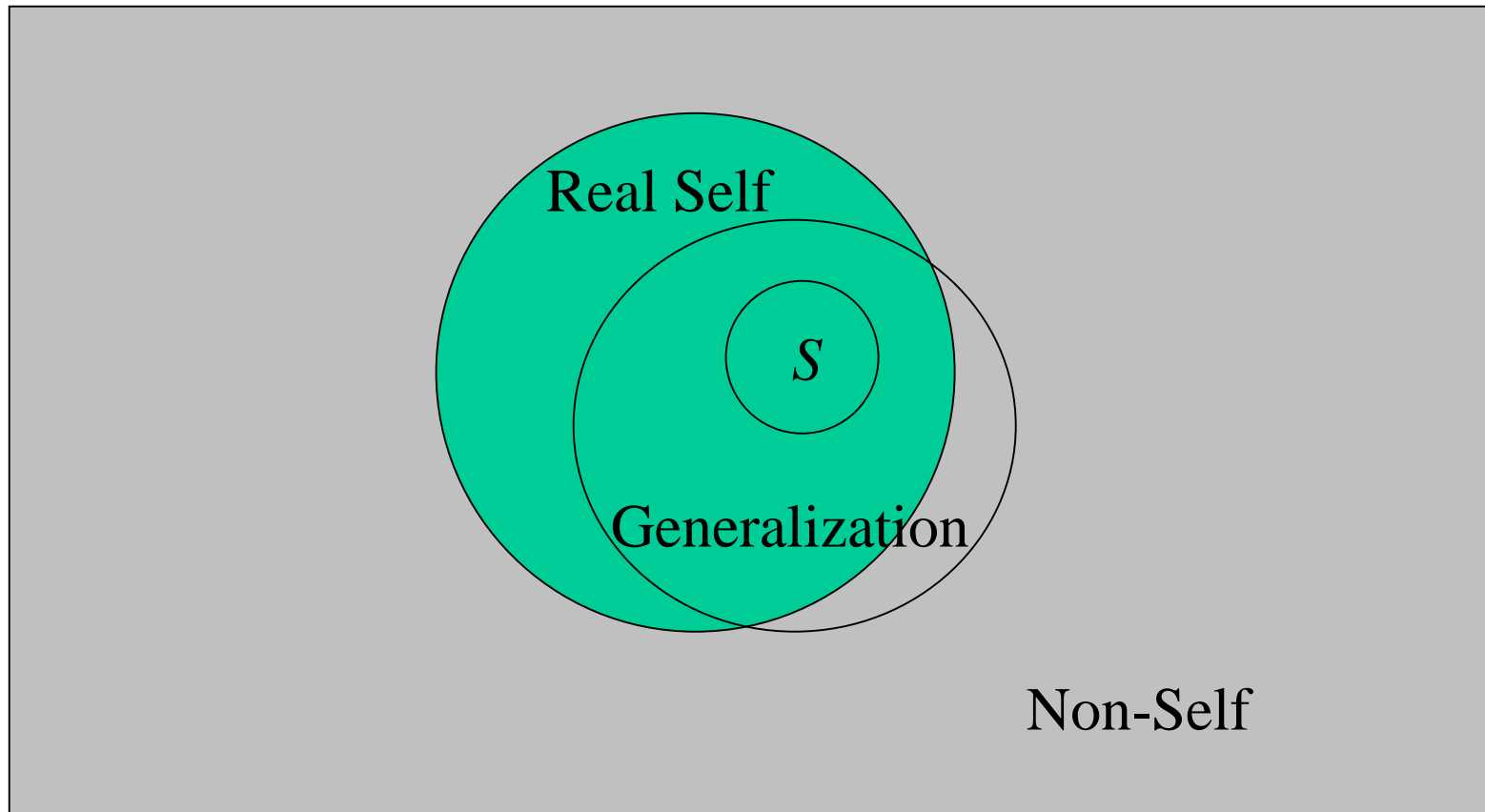


Acknowledgments

- National Science Foundation (grants CDA-9503064, and ANIR-9986555)
- Office of Naval Research (grant N00014-99-1-0417)
- Defense Advanced Projects Agency (grant AGR F30602-00-2-0584)
- Intel Corporation
- Santa Fe Institute.

URL: www.cs.unm.edu/~immsec/

Generalization



- Given a sample S what strings are encompassed in the generalization?