2012-08-02

Approaches to Event Prediction in Complex Environments

Tan, Terence

Monterey, California: Naval Postgraduate School.

http://hdl.handle.net/10945/44407

Approaches to Event Prediction in Complex Environments

Terence Tan (PhD Candidate)

Advisors:

Prof Christian Darken, PhD

Prof Neil Rowe , PhD

Prof Arnold Buss , PhD

Prof Ralucca Gera , PhD

Prof John Hiles

Scope of Presentation

- What is Relational Time Series?
- Previous Approaches
- New Learning and Prediction Approaches
- Conclusions

Network Intrusion Detection Alerts

12/02/11-17:32:21.984133	2924	NETBIOS SMB-DS repeated logon failure	TCP	78.45.215.210	63.205.26.80

Network Intrusion Detection Alerts

12/02/11-17:32:21.984133	2924	NETBIOS SMB-DS repeated logon failure	TCP	78.45.215.210	63.205.26.80
12/02/11-17:32:24.712867	2924	NETBIOS SMB-DS repeated logon failure	TCP	78.45.215.210	63.205.26.80

Network Intrusion Detection Alerts

12/02/11-17:32:21.984133	2924	NETBIOS SMB-DS repeated logon failure	TCP	78.45.215.210	63.205.26.80
12/02/11-17:32:24.712867	2924	NETBIOS SMB-DS repeated logon failure	TCP	78.45.215.210	63.205.26.80
12/02/11-18:50:13.575037	648	SHELLCODE x86 NOOP	TCP	84.0.158.110	63.205.26.70

Network Intrusion Detection Alerts

12/02/11-17:32:21.984133	2924	NETBIOS SMB-DS repeated logon failure	TCP	78.45.215.210	63.205.26.80
12/02/11-17:32:24.712867	2924	NETBIOS SMB-DS repeated logon failure	TCP	78.45.215.210	63.205.26.80
12/02/11-18:50:13.575037	648	SHELLCODE x86 NOOP	TCP	84.0.158.110	63.205.26.70
12/02/11-18:50:13.575356	648	SHELLCODE x86 NOOP	TCP	84.0.158.110	63.205.26.70

Time Series of Network Intrusion Detection Alerts

12/02/11-17:32:21.984133	2924	NETBIOS SMB-DS repeated logon failure	TCP	78.45.215.210	63.205.26.80
12/02/11-17:32:24.712867	2924	NETBIOS SMB-DS repeated logon failure	TCP	78.45.215.210	63.205.26.80
12/02/11-18:50:13.575037	648	SHELLCODE x86 NOOP	TCP	84.0.158.110	63.205.26.70
12/02/11-18:50:13.575356	648	SHELLCODE x86 NOOP	TCP	84.0.158.110	63.205.26.70
12/02/11-18:50:13.575356	3397	NETBIOS DCERPC NCACN-IP-TCP	TCP	84.0.158.110	63.205.26.70
12/02/11-18:50:15.443929	648	SHELLCODE x86 NOOP	TCP	84.0.158.110	63.205.26.73
12/02/11-18:50:15.444255	648	SHELLCODE x86 NOOP	TCP	84.0.158.110	63.205.26.73
12/02/11-18:50:15.444255	3397	NETBIOS DCERPC NCACN-IP-TCP	TCP	84.0.158.110	63.205.26.73
12/02/11-18:50:19.048303	648	SHELLCODE x86 NOOP	TCP	84.0.158.110	63.205.26.77
12/02/11-18:50:19.048624	648	SHELLCODE x86 NOOP	TCP	84.0.158.110	63.205.26.77
12/02/11-18:50:19.048624	3397	NETBIOS DCERPC NCACN-IP-TCP	TCP	84.0.158.110	63.205.26.77
12/02/11-18:50:20.346232	648	SHELLCODE x86 NOOP	TCP	84.0.158.110	63.205.26.74
12/02/11-18:50:22.656974	648	SHELLCODE x86 NOOP	TCP	84.0.158.110	63.205.26.79
12/02/11-18:50:22.657291	648	SHELLCODE x86 NOOP	TCP	84.0.158.110	63.205.26.79
12/02/11-18:50:22.657291	3397	NETBIOS DCERPC NCACN-IP-TCP	TCP	84.0.158.110	63.205.26.79
12/02/11-19:12:38.913940	384	ICMP PING	ICMP	66.235.66.233	63.205.26.80
12/02/11-19:12:38.914642	408	ICMP Echo Reply	ICMP	63.205.26.80	66.235.66.233
12/02/11-19:12:38.959461	384	ICMP PING	ICMP	66.235.66.233	63.205.26.80
12/02/11-19:12:38.959672	408	ICMP Echo Reply	ICMP	63.205.26.80	66.235.66.233

Relational Time Series: Time Series of Relational Atoms

- 0.0, lookA(spock84)
- 0.0, place+(Paperville3)
- 0.0, location+(pitchfork74, Paperville3)
- 0.0, pitchfork+(pitchfork74)
- 0.0, location+(spock84, Paperville3)
- 0.0, spock+(spock84)
- 2.75, getA(pitchfork74, spock84)
- 2.75, getE(spock84, pitchfork74)
- 2.75, location-(pitchfork74, Paperville3)
- 2.75, location+(pitchfork74, spock84)
- 5.5, wA(spock84)
- 5.5, goE(spock84, west)
- 5.5, location-(spock84, Paperville3)
- 5.5, spock-(spock84)
- 5.5, place-(Paperville3)

$P = (t, r(c_1, c_2, \dots c_n))$
where
P: percept
t: time
r: relation
c_x: constant

Characteristics

- No Background knowledge
  - Eg. In a unknown domain, we do not know the behaviors of any entity
- Relational Atoms
  - Multi-dimension proposition
- High variability in predicates & constants
  - Too many to predefine
- Moving Context
  - Needs online Learning

Possible Approaches

- Approaches
  - Production Rules
  - Finite State Machines
  - Bayesian Network
  - Markov Chain
  - Statistical Relational Learning
- Recent Interest in IDS Alerts Predictions
  - 2011 Nexat a history-based approach to predict attacker actions
  - 2011 A Novel Probabilistic Matching Algorithm for Multi-Stage Attack Forecast
  - 2010 Multi stage attack Detection system for Network Administrators using Data Mining (UTN, Oak Ridge NL)
  - 2008 Alert Fusion Based on Cluster and Correlation
  - 2007 Using Network Attack Graph to Predict the Future Attacks
  - 2007 Discovering Novel Multistage Attack Strategies

Situation Learning

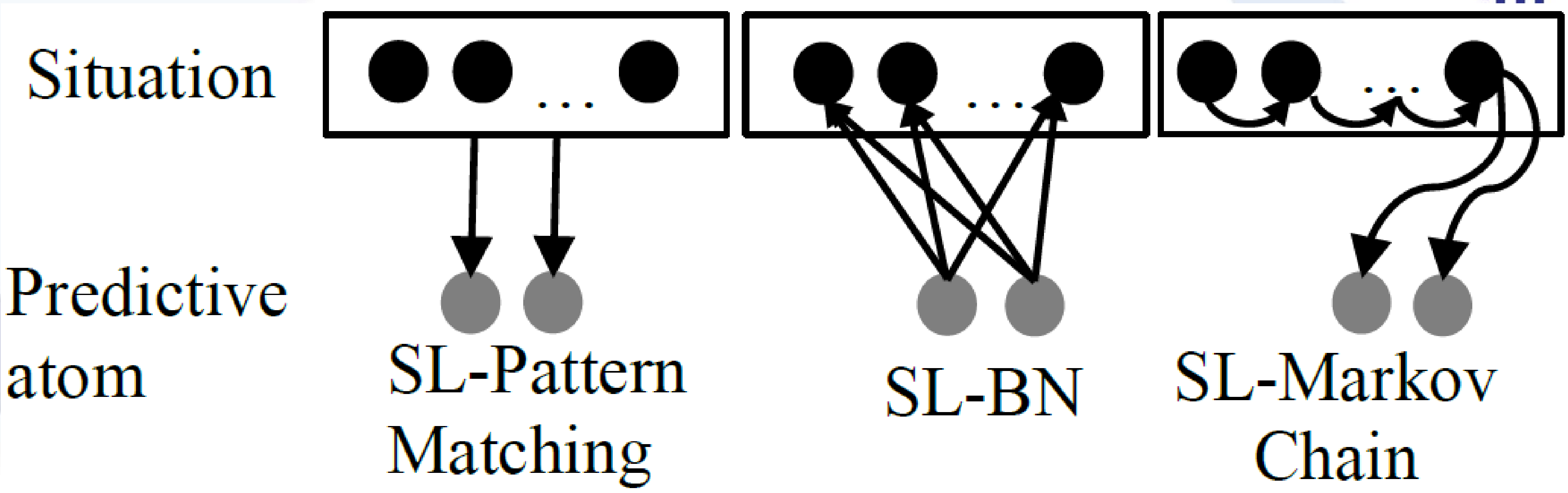
- Situation Learning (Darken, 2005)
  - A sliding time window identifies "Situations"
  - Forms a simple lookup table
  - Able to model trending and high variability

- 0.0, lookA(spock84)
- 0.0, place+(Paperville3)
- 0.0, location+(pitchfork74, Paperville3)
- 0.0, pitchfork+(pitchfork74)
- 0.0, location+(spock84, Paperville3)
- 0.0, spock+(spock84)
- 2.75, getA(pitchfork74, spock84)
- 2.75, getE(spock84, pitchfork74)
- 2.75, location-(pitchfork74, Paperville3)
- 2.75, location+(pitchfork74, spock84)
- 5.5, wA(spock84)
- 5.5, goE(spock84, west)
- 5.5, location-(spock84, Paperville3)
- 5.5, spock-(spock84)
- 5.5, place-(Paperville3)

Δt

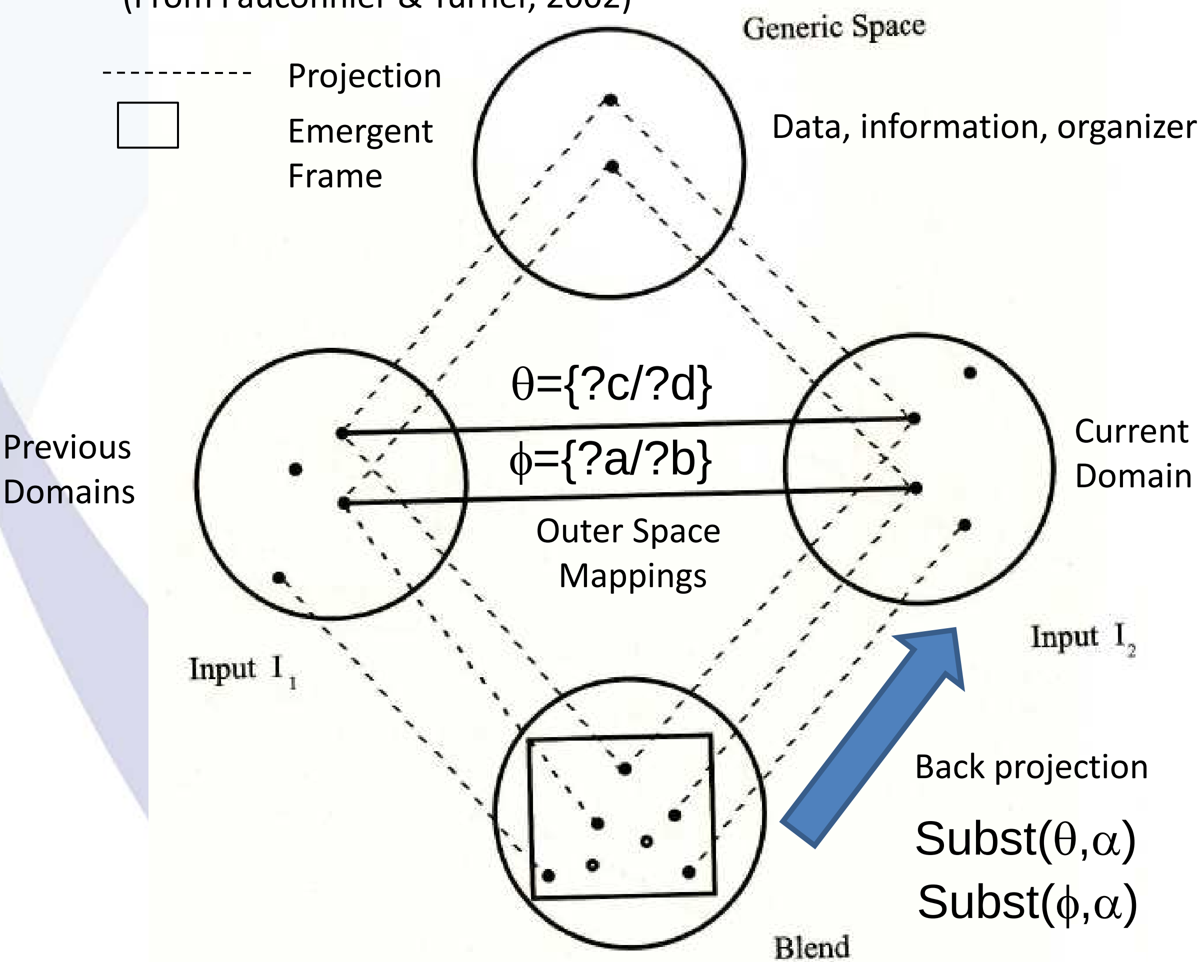
Predictive atom

Situation Learning (SL) + Current Approaches



Conceptual Blending

(From Fauconnier & Turner, 2002)



Constitution Principles
Vital Relation Mapping
Construct Generic Space
Composition
Completion
Elaboration
Back Projection

Optimality Principles
Compression
Topology
Pattern Completion
Integration
Promoting Vital Relation
Web
Unpacking
relevance

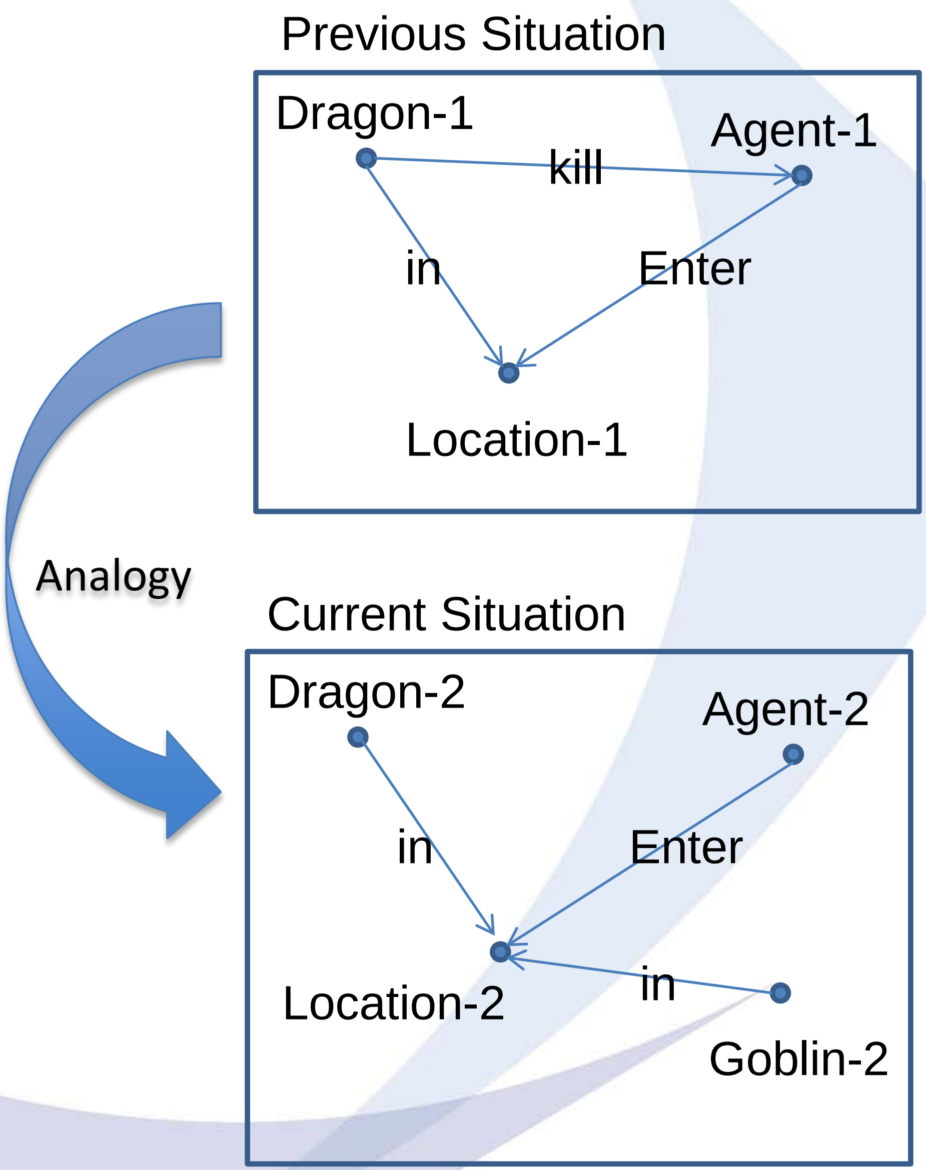
Vital Relation
Change, Cause-Effect, Time, Space, Identity, Change, Uniqueness, Part-Whole, Representation, Role, Analogy, Disanalogy, Property, Similarity, Category, and Intentionality

Single Scope Blending (SSB)

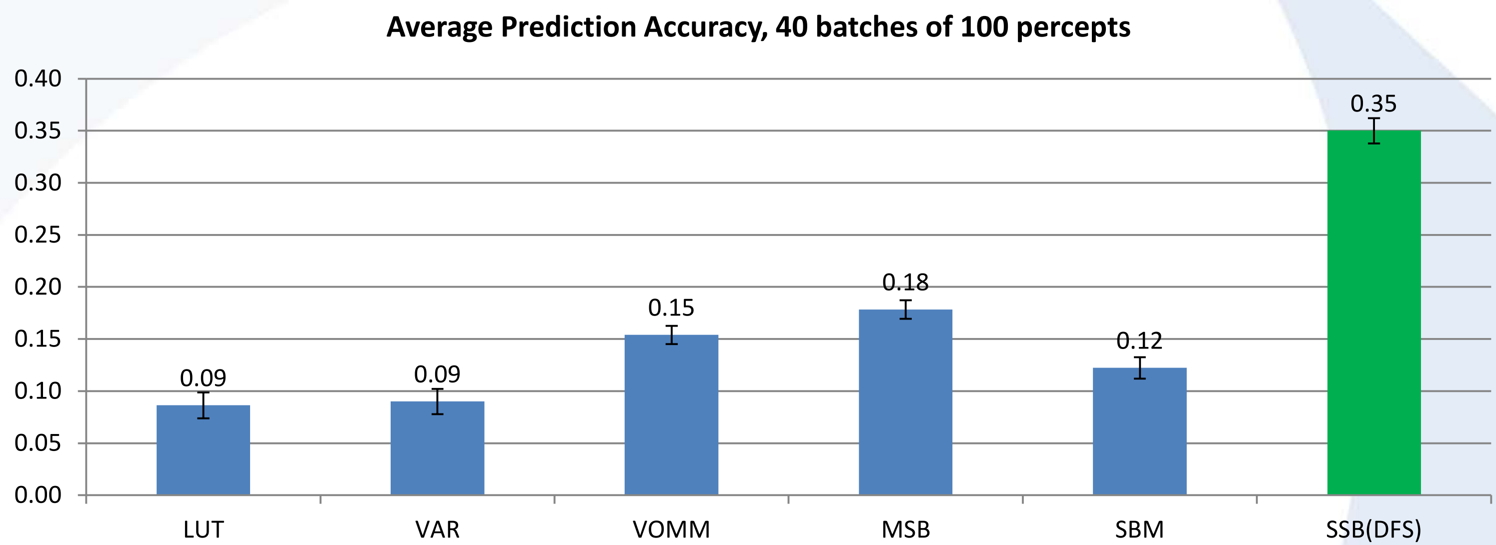
- Dragon-1 in Location-1
- Agent-1 enter Location-1
- Dragon-1 Kill Agent-1
- …
- …
- Goblin-2 in location-2
- Dragon-2 in location-2
- Agent-2 enter location-2
- ?

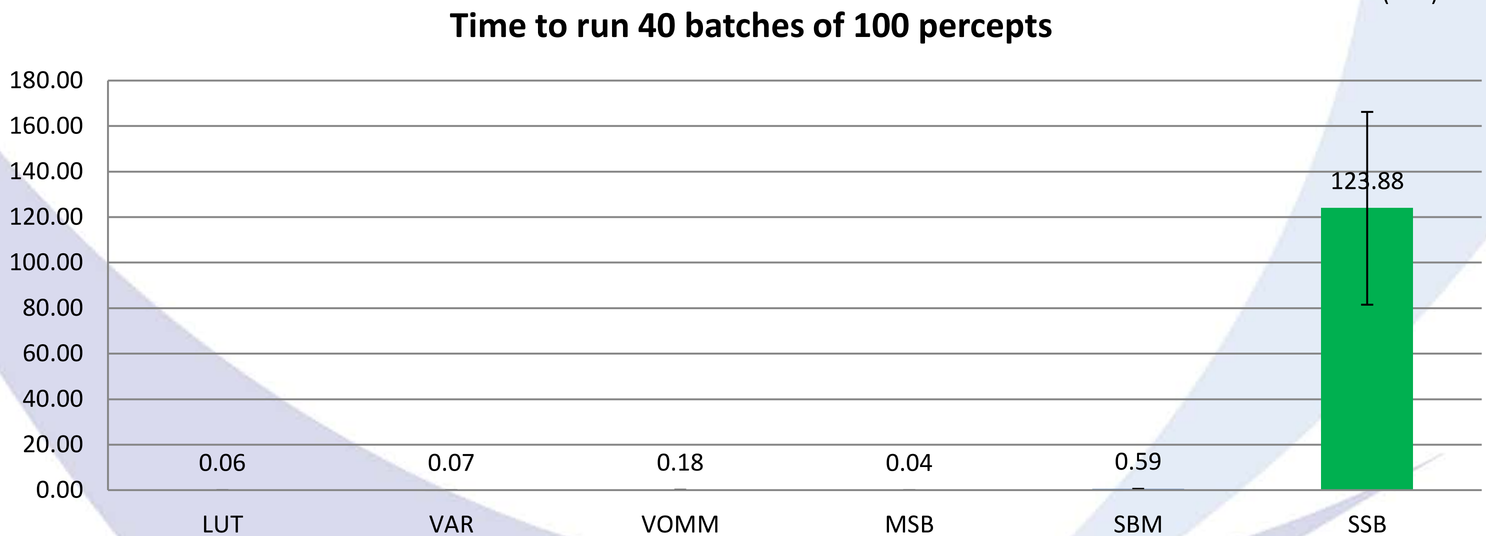
One Possible Substitution:
Dragon-1 to Goblin-2
Agent-1 to Agent-2

Prediction: Goblin-2 Kill Agent-2



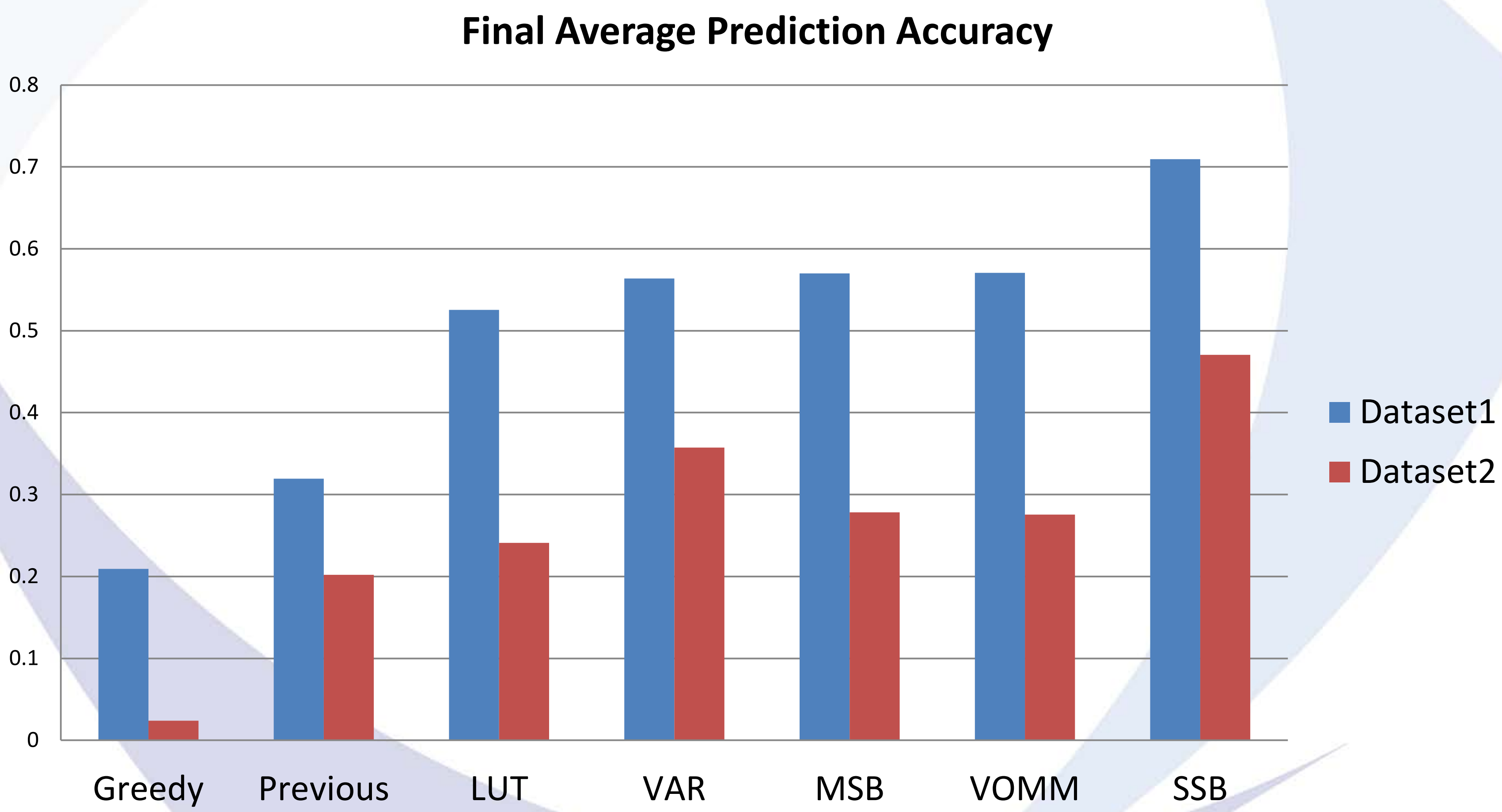
Prediction Accuracies from a Agent Simulator





What is Relational Time Series?
Previous Approaches
New Learning and Prediction Approaches
Conclusions

Network Intrusion Alerts Predictions



Why is SSB Better?

- Dataset
  - 6482 alerts
  - 1590 unique alerts
- Detection Rate

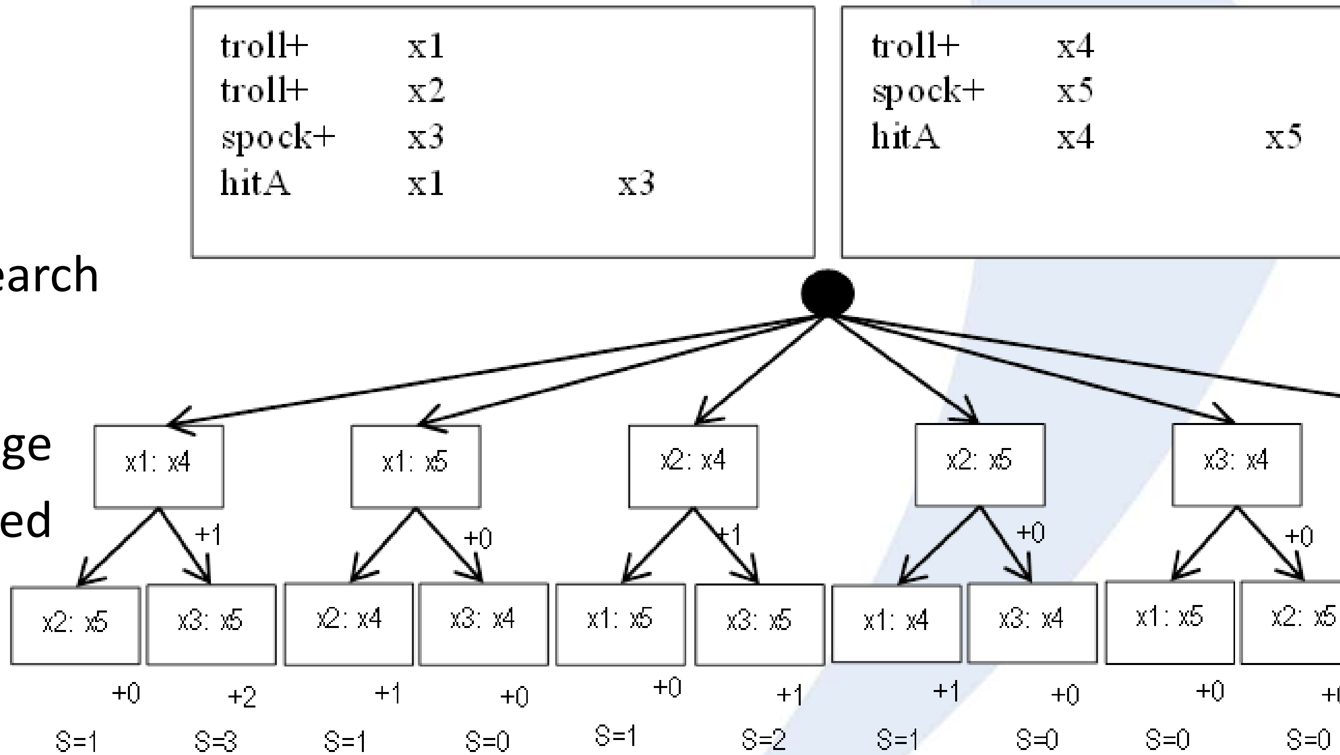
	SSB	MSB	VOMM
Unique Alert Detected	947	379	375
%	59.56%	23.84%	23.58%

- Effect of Frequency on Detection Rate

Frequency	Number of Alerts	SSB Detects	MSB detects	VOMM detects
1	643	163	0	0
2	751	621	230	242
3	52	34	27	14
4	88	80	77	74
5	5	0	0	0
6	11	10	8	8
7	3	3	1	1
8	2	2	2	2
9	4	4	3	3
10	3	3	3	3

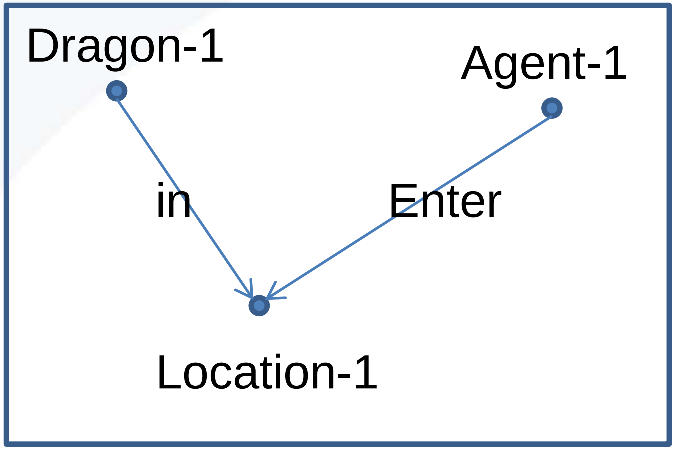
Complexity Reduction: From Exponential to near Linear

- Default Method: Backtracking
  - Subgraph Isomorphism
  - NP-Complete
- Improvements
  - Greedy ASTAR
  - Attention Based Search
- Results
  - Accuracy: No Change
  - Complexity: Reduced

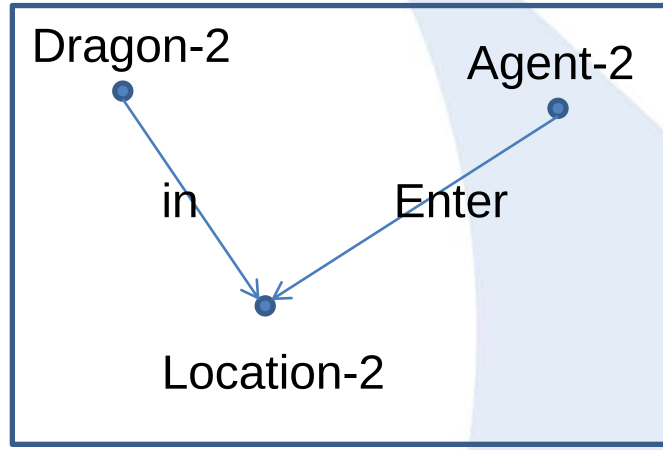


Attention Based Search

Previous Situation



Current Situation

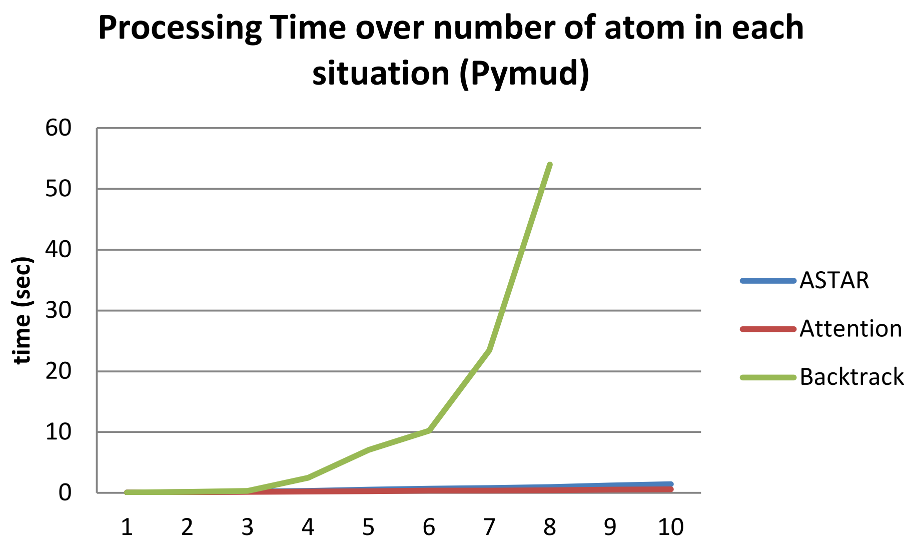


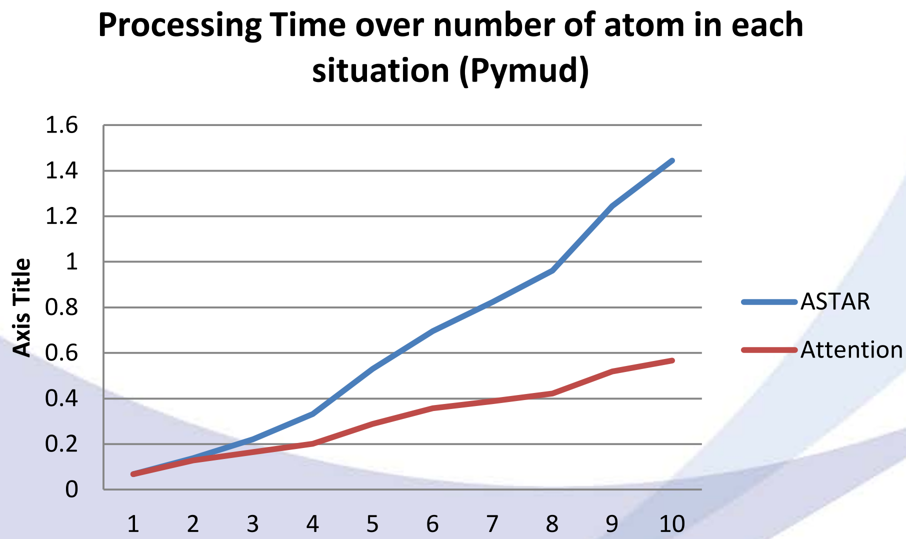
nodes	In Degree	Out Degree	Type
Dragon – 1	0	1	D
Agent – 1	0	1	A
Location – 1	2	0	L
Dragon – 2	0	1	D
Agent – 2	0	1	A
Location – 2	2	0	L

node1	node2	Difference
Dragon - 1	Dragon - 2	[1, 0, 1, 1, 0]
	Agent - 2	[0, 0, 1, 1, 0]
	Location - 2	[0, 0, 0, 0, -3]
Agent - 1	Dragon - 2	[0, 0, 1, 1, 0]
	Agent - 2	[1, 0, 1, 1, 0]
	Location - 2	[0, 0, 0, 0, -3]
Location - 1	Dragon - 2	[0, 0, 0, 0, -3]
	Agent - 2	[0, 0, 0, 0, -3]
	Location - 2	[1, 0, 1, 1, 0]

score = [Type, ExactNameMatch, BothExactDegreeMatch, AtLeastOneDegreeMatch, DegreeDiff]

Scalability Test





Conclusions

- Single Scope Blending Prediction Approach predicts better
- Reduces NP-Complete complexity to Linear through Greedy ASTAR and Attention based search

Thank you