



Network Based Hard/Soft Information Fusion

Stochastic Graph Analytics

Geoff Gross, Kedar Sambhoos, Rakesh Nagi (PI)
Tel. (716) 645-3471, Email: gagross@buffalo.edu

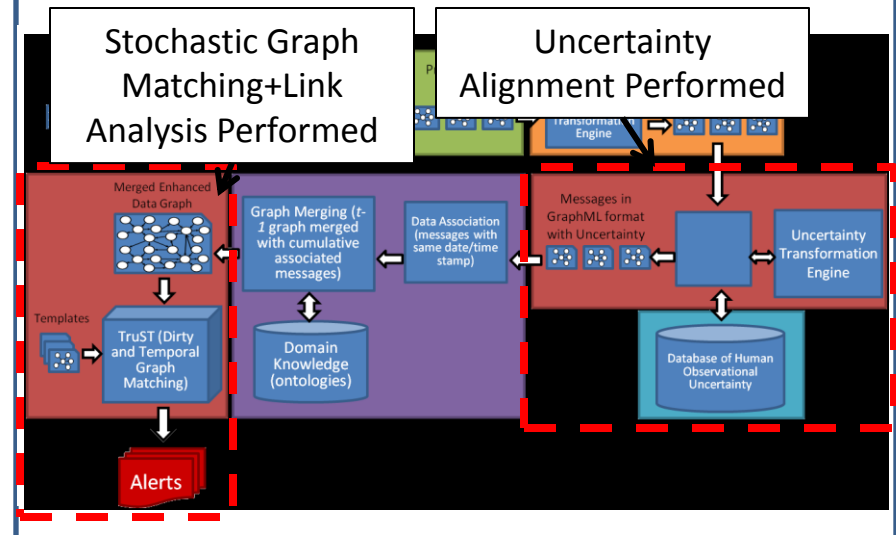


Objectives

- Represent soft data uncertainties in a unified framework
- Utilize uncertainty in situation assessment process
 - Indicate level of uncertainty in the potential situation of interest or hypothesis
- Implement incremental methods to allow for maintenance of real-time situation awareness
- Enable intelligent querying methods within graph matching algorithm
- Develop scalable link analysis solution for identifying connections between multiple entities of interest

Scientific/Technical Approach

- Incremental graph matching result updating developed to produce result better than or equal to batch Truncated Search Tree (TruST) run
- Graph matching algorithm extended for intelligent templates
 - AND/OR templates (decompose PIRs to avoid recalculations when possible)
- Map/reduce multi-entity of interest link analysis algorithm developed



Accomplishments

- Incremental graph matching methodology produces results better than or equal to batch run with significant runtime benefit
- AND/OR templates enable many scenarios which previously required multiple templates to be located in a single graph matching run
- Multi-entity of interest link analysis algorithm implemented within Hadoop framework



Main Scientific/Technical Accomplishments



Situation Assessment: Stochastic Graph Analytics



Motivation

- Analysts respond to the existence of many different threats with some common elements
 - Rigid template structure requires many templates with similar elements, resulting in largely redundant calculations
 - Must develop flexible (AND/OR) templates to reduce overall number of graph matching executions
 - AND/OR templates decompose PIRs, avoiding recalculations on common graph elements
- Analysts are often interested in loose (many-hop) connections between entities in evidential data
 - e.g., distant familial relationships
 - Requires scalable link analysis solution to handle the size and complexity of such queries



Main Scientific/Technical Accomplishments

Situation Assessment: Stochastic Graph Analytics



Scientific/Technical Accomplishments

- Developed incremental graph matching approach which produces results better than or equal to batch graph matching approach
- Enabled the use of templates with both AND and OR relationships on template graph nodes and edges
- Implemented multi-entity of interest link analysis algorithm implemented within Hadoop framework



Main Scientific/Technical Accomplishments

Situation Assessment: Stochastic Graph Analytics



Years 1 and 2

- Developed framework for representing uncertainties in soft data
 - Utilized uncertainties through situation assessment processes
- Implemented incremental graph matching approach for the maintenance of continuous situational awareness

Year 3 Accomplishments

- Performed extensive numerical testing of methods implemented in Years 1 and 2 (stochastic/incremental graph matching)
- Developed map/reduce link analysis algorithm to identify paths between multiple entities of interest



Project Statistics and Summary

Situation Assessment: Stochastic Graph Matching



Degree Awarded

- M.S. – Geoff Gross
- Ph.D – Geoff Gross (Expected June 2013)

Students Supported

- Number of Graduate Students: 1

Publications

- Journal papers – 2 in review, 1 in progress
- Conference papers – 1

Technology Transitions

- Graph Matching
 - Discussions with ARL-APG (Mark Thomas)
 - Support to Dr. Kirk Ogaard (postdoctoral researcher – ARL; Sue Kase)
- Link Analysis
 - Discussion with A2SF (Milan Patel)



Publication List

Situation Assessment: Stochastic Graph Matching



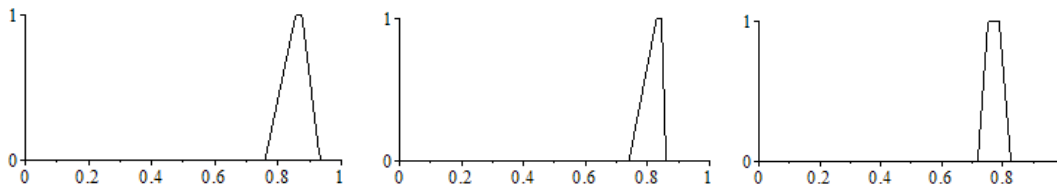
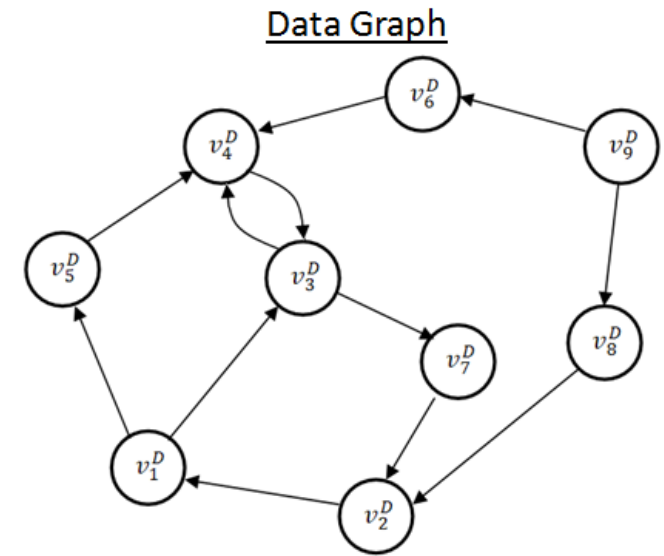
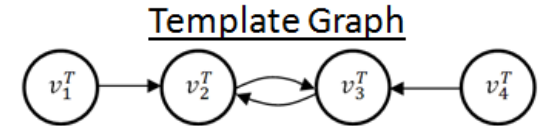
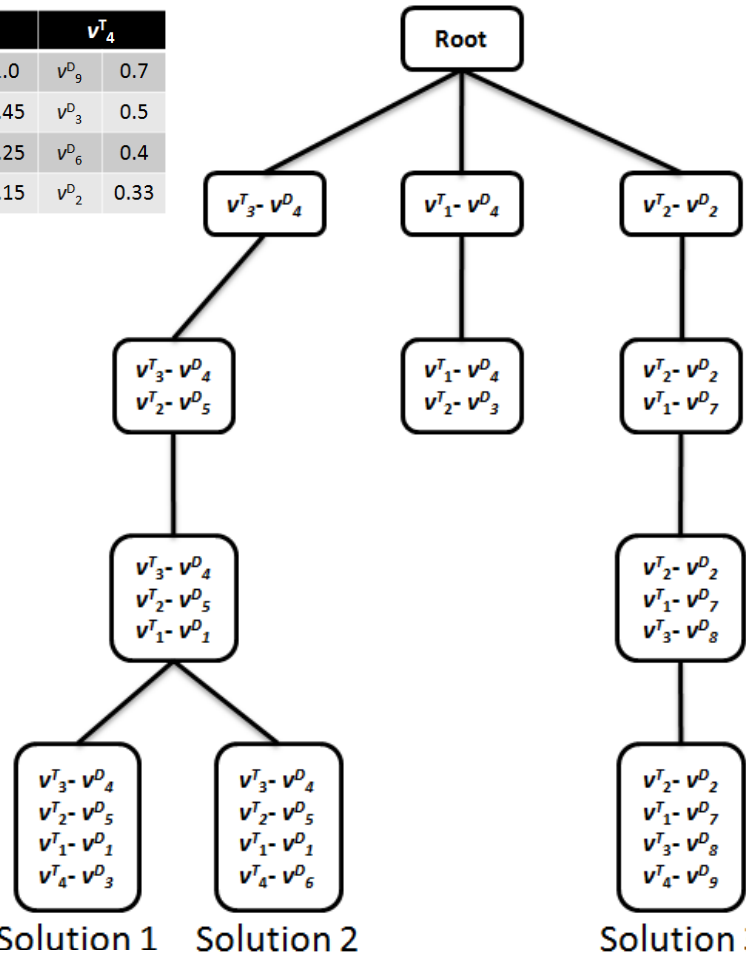
- Gross, G., Nagi, R., Graph Matching Search Tree Precedence and AND/OR Template Matching, (In Progress)
- Gross, G., Nagi, R., and Sambhoos, K., A Fuzzy Graph Matching Approach in Intelligence Analysis and Maintenance of Continuous Situational Awareness, *Journal of Information Fusion*, (Revisions Submitted, September 2012)
- Gross, G., Nagi, R., Sambhoos, K., Schlegel, D., Shapiro, S., Tauer, G., Towards Hard+Soft Data Fusion: Processing Architecture and Implementation for the Joint Fusion and Analysis of Hard and Soft Intelligence Data, *The 15th International Conference on Information Fusion*, Singapore, July 2012
- Jenkins, M., Gross, G., Bisantz, A. M., and Nagi, R., Towards Context Aware Data Fusion: Modeling and Integration of Situationally Qualified Human Observations into a Fusion Process for Intelligence Analysis, *Journal of Information Fusion*, (In Revision, September 2012)



Technical Approach

Graph Matching Search Tree Execution

	v_1^T	v_2^T	v_3^T	v_4^T
v_4^D	1.0	1.0	1.0	0.7
v_1^D	0.8	0.85	0.45	0.5
v_7^D	0.8	0.7	0.25	0.4
v_3^D	0.4	0.2	0.15	0.33





Incremental Result Updating

Reconstruction Point Identification

Reconstruction Point – a point in the search tree at which lower search tree levels must be reformed

- Caused by additions/deletions in the data graph

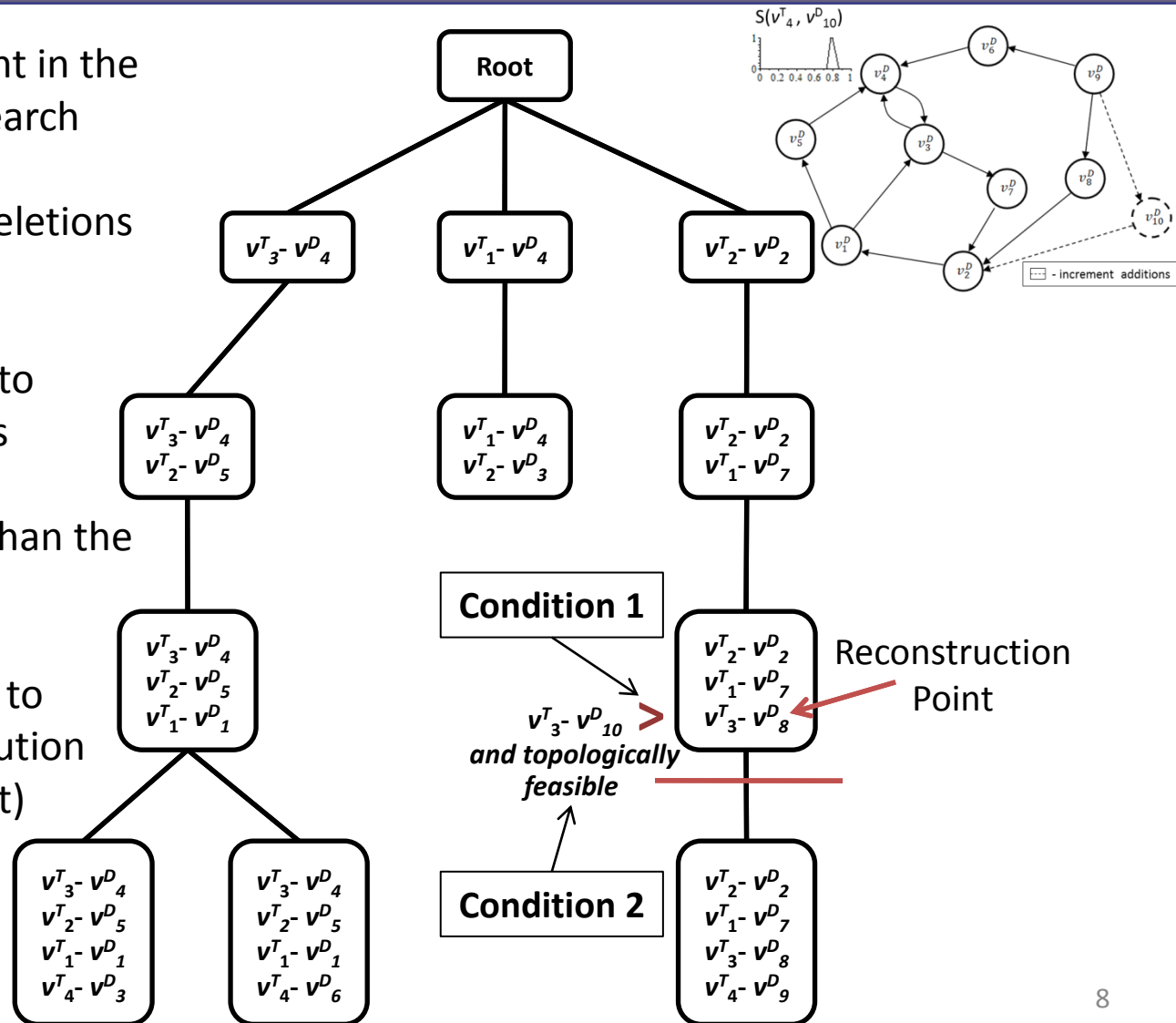
Two conditions must be met to identify reconstruction points

Condition 1

- A changed score is greater than the score added at that level

Condition 2

- A changed score is adjacent to existing partial solution (solution prior to reconstruction point)





Technical Approach



Stochastic Graph Analytics – Numerical Testing



- Graph analytic techniques of stochastic and incremental graph matching went through extensive numerical testing
- Pilot study run to provide algorithmic parameters for testing
- Verified solution quality and runtime versus commercial solver (CPLEX)
 - see paper below for full result details

k_0 Settings

Data Graph Size

		500	1000	2000	3000	4000	5000
Template	3	300	300	500	600	600	600
Graph	4	300	600	700	1100	1100	1100
Size	5	400	600	700	1100	1100	1100
(Nodes)	6	400	600	900	1100	1100	1400

β Settings

Data Graph Size

		500	1000	2000	3000	4000	5000
Template	3	600	600	1250	2250	2250	2250
Graph	4	750	2250	3000	6000	6000	6000
Size	5	1250	2250	3000	6500	6500	6500
(Nodes)	6	1250	2500	4000	6500	6500	7500



Technical Approach

Stochastic Graph Matching

Overview and Solution Quality Comparison



- CPLEX requires 42.7 times more time to identify top 10 results than graph matching execution
 - Time averaged for data graphs from 500-5,000 nodes
 - Runtime benefit multiplier increases in size of data graph
- Minimal optimality gap under all experimental conditions
 - 17 of 24 experimental conditions experience no gap (all optimal solutions obtained)

Average Optimality Gap (10 Results)

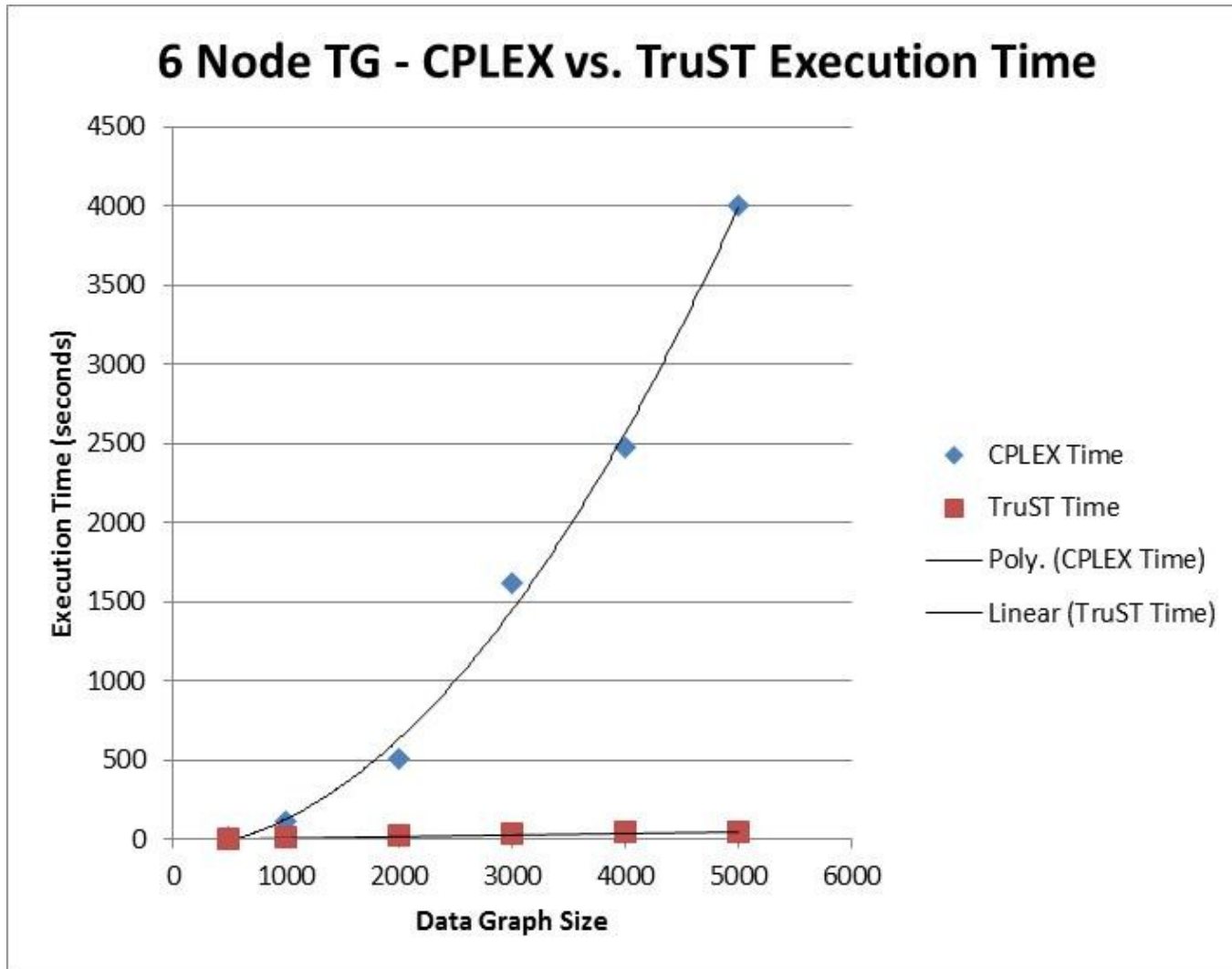
Data Graph Node Count

	500	1000	2000	3000	4000	5000	
TG Node Count	3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
4	0.04%	0.00%	0.01%	0.00%	0.18%	0.34%	
5	0.00%	0.00%	0.00%	0.00%	0.00%	0.11%	
6	0.00%	0.00%	0.00%	0.00%	0.23%	0.07%	



Technical Approach

Stochastic Graph Matching Execution Time Comparison





Technical Approach



Incremental vs. Batch Stochastic Graph Matching Execution Time Comparison



- Incremental results indicate significant speedup over batch graph matching algorithm
 - Speedup with search tree reconstruction simulates environment where search tree cannot be maintained in memory
 - In this case search tree must be rebuilt from previous search tree leaf nodes

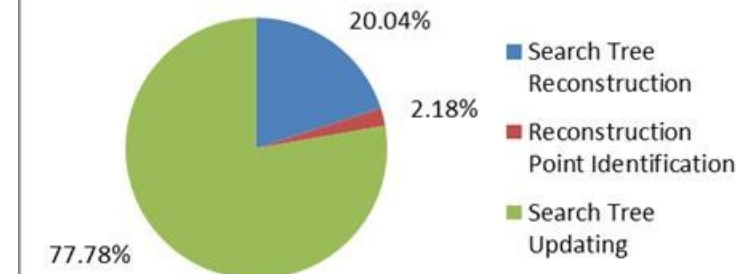
Template Graph Node Count

	3	4	5	6
Incremental Speedup with Search Tree Reconstruction	36.2	44.5	33.9	36.9
Incremental Speedup without Search Tree Reconstruction	46.4	60.8	41.5	42.9

Incremental Speedup with Search Tree Reconstruction

Incremental Speedup without Search Tree Reconstruction

Incremental Algorithm Runtime Breakdown

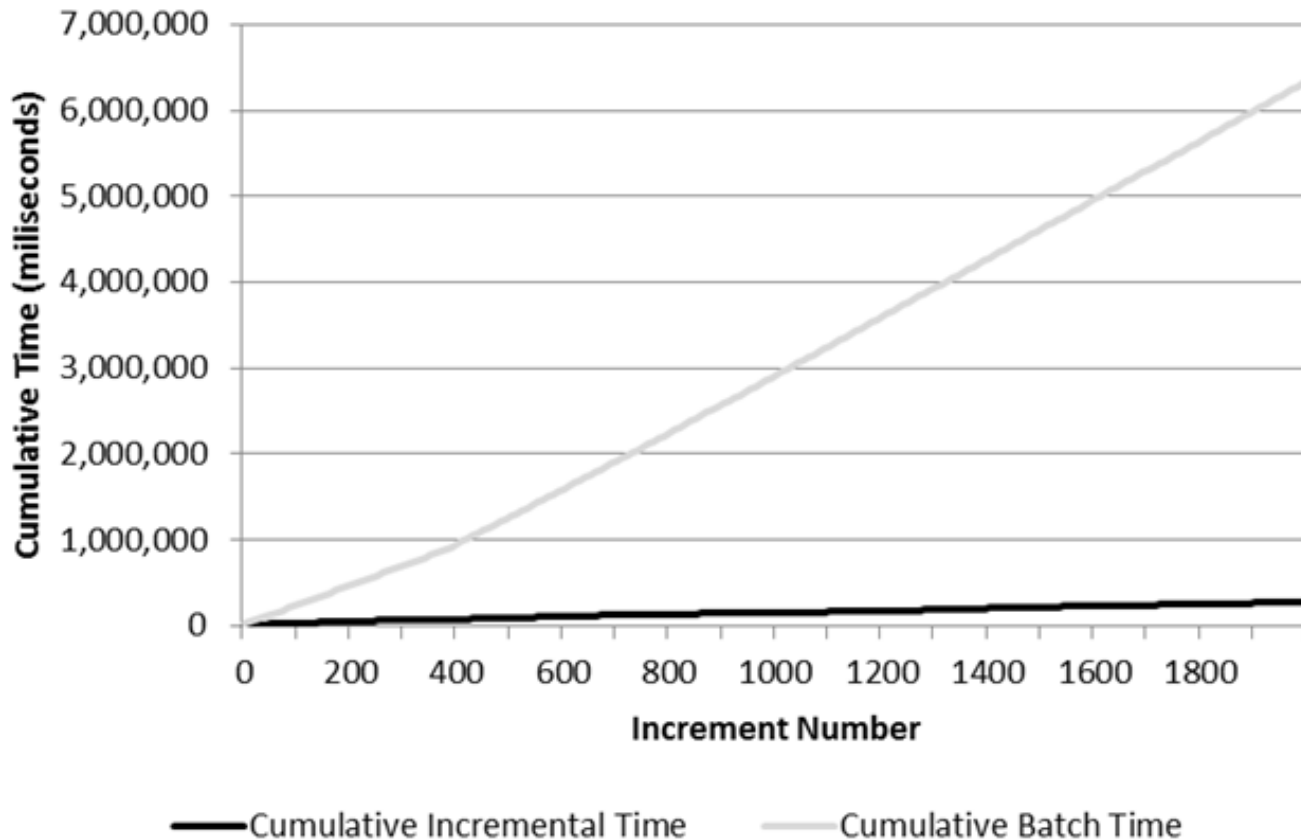




Technical Approach

Incremental vs. Batch Stochastic Graph Matching Cumulative Execution Time Comparison

6 Node TG Cumulative Runtime - Incremental vs. Batch Graph Matching Execution





Technical Approach

Link Analysis

(Collaboration with A2SF)



Task Description

- Given two (or more) entities of interest (EOI), identify all paths connecting these nodes within the global graph
 - Termination Criteria
 - Number of hops
 - Number of solutions
 - Runtime
 - Analyst interrupt



Technical Approach

Link Analysis

(Collaboration with A2SF)



Map-Reduce Implementation Overview

1. Identify instances of the entities (root nodes) of interest within the global graph
2. Branch from the existing frontier (root nodes only for first iteration)
3. Determine overlap of newly reached nodes with paths from other root nodes
 - Output connected paths
4. Check termination criteria, if not met return to Step 2



Software Implementations

Uncertainty Alignment Space Builder

(Proposed for ARL Infrastructure Program)



Uncertainty Alignment Editor

File Views

Entity Values

New Delete Edit

Building
Person

Attributes

New Delete Edit

Name	Range Low	Range High
numerosity	1.0	99.0
size	1.0	5.0

Uncertainty Alignment Space

New Delete Edit

Range Low: 1.0, Range High: 4.0
Range Low: 11.0, Range High: 15.0
Range Low: 16.0, Range High: 37.0
Range Low: 38.0, Range High: 99.0
Range Low: 5.0, Range High: 5.0
Range Low: 6.0, Range High: 10.0

Observer Characteristics

New Delete Edit

numerosity training

Environment Characteristics

New Delete Edit

dark out

Uncertainty Model

New Delete Edit

Bias Model

Bias: 0

Variance Model

Distance to - LB: -1.0, LC: 0.0, HC: 0.0, HB: 1.0

Distribution Family: fuzzy number

Uncertainty Alignment Condition Selection

Entity: Building

Observer Characteristics: numerosity

Environment Characteristics:

Observed Value: _____

View Uncertainty Alignment Space

Uncertainty Representation

Y-axis: 0.0 to 1.0

X-axis: 0.0 to 1.0

Uncertainty Representation



Software Implementations

Uncertainty Alignment

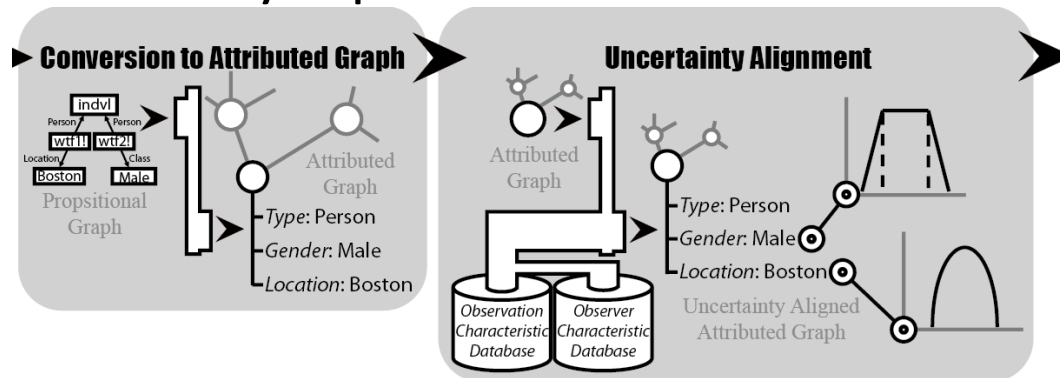


- **Uncertainty Alignment Input**

- **Uncertainty Alignment Database** – models of human observation and influencing contextual factors
- **Observer Database** – data about potential observers known to be relevant to observational accuracy/precision
- **Un-aligned Message Graphs** – attributed graphs which must be aligned

- **Uncertainty Alignment Output**

- **Uncertainty Aligned Message graphs** – attributed graphs with uncertainty representations at the attribute level





Software Implementations

Association/Graph Matching Scoring

- **Scoring Input**

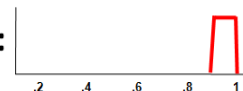
- Uncertainty aligned message/cumulative graphs for association and graph matching respectively
- Template graph(s) for graph matching scoring

- **Scoring Output**

- Similarity scores for use in association or graph matching process

Attribute	Target Person	Observed Person	Attribute Similarity
Height	6.08'	tall	
Age	36-39	42	
Weight	fuzzy{200, 215, 230}	210-220	

Node-Node Similarity:





Software Implementations

AND/OR Template Graph Builder



TruST Results Viewer

File Graphs Views Uncertainty Alignment Association

Graph View

Precedence View

Attribute View AND/OR View

OR Elements

OR Element ID	Nodes	Edges	OR Relation on Edges?
1	ID: 4, ID: 3, ID: 2	3_4, 2_3	yes
2	ID: 1	2_1	yes
3	ID: 6, ID: 5	4_5, 5_6	yes
4	ID: 8, ID: 7	4_7, 7_8	yes
5	ID: 9, ID: 7	7_9, 4_7	yes

OR Sets

OR Set ID	OR Elements
1	1
2	3, 2, 5, 4

AND Edges

AND Edge ID	OR Set 1	OR Set 2
1	1	2



Software Implementations

Graph Matching Execution



TruST Results Viewer

File Graphs Views Uncertainty Alignment Association

Trust Execution View

Parameters

Run greedy heuristic assignment algorithm	<input checked="" type="checkbox"/>
t	0
p	1
alpha	0.5
node-node weight	1.0
ko	200
ki	100
beta	500
delta	10000
TruST output file name	GMOutput.xml
Set the template graph file	C:\Users\Geoff\workspace\TrustViewer\Year 3 Demo\TGs\Line02_IEDPossession.xml.layout
Set the data graph file	C:\Users\Geoff\workspace\TrustViewer\Year 3 Demo\CDG_GT_2012-09-17.xml
Set the entity-attribute weighting file	
Set the TruST output file directory	C:\Users\Geoff\workspace\TrustViewer\Year 3 Demo\GM_Out
Set the plain text message containing directory.	C:\Users\Geoff\workspace\TrustViewer\Year 3 Demo\Messages
Set the node to node score file (if one already exists)	
Set the edge to edge score file (if one already exists)	

Output

```
Nodes : 3 Edges : 2 File size: 788 Bytes
Creating the graph representation for C:\Users\Geoff\workspace\TrustViewer\Year 3 Demo\TGs\Line02_IEDPossession.xml consumed 20 milliseconds.

Nodes : 167 Edges : 224 File size: 314888 Bytes
Creating the graph representation for C:\Users\Geoff\workspace\TrustViewer\Year 3 Demo\CDG_GT_2012-09-17.xml consumed 29 milliseconds.

Parsing of the input consumed 125 milliseconds.

Computation of one hop neighborhood scores consumed 40 milliseconds.

All scoring consumed 40 milliseconds.

Branch to level: 1, with: 200 branch roots.
Branch to level: 2, with: 386 branch roots.
Branch to level: 3, with: 338 branch roots.

The truncated search algorithm consumed 883 milliseconds.

The execution of Trust for C:\Users\Geoff\workspace\TrustViewer\Year 3 Demo\GM_Out\GMOutput.xml consumed 1075 milliseconds.
```

Stop TruST Execution



Software Implementations

Graph Matching Results Viewer



TruST Results Viewer

File Graphs Views Uncertainty Alignment Association

Results

Result Number	Overall Score
1	1.0
2	0.634921
3	0.634921
4	0.634921
5	0.634921
6	0.634921
7	0.634921
8	0.634921
9	0.634921
10	0.634921
11	0.634921
12	0.634921

Overall score: 1.0
 Weighted node score: 0.666667
 Node score: 1.0
 Number of matched nodes: 3
 Sum of node match values: 3.0
 Weighted edge score: 0.333333
 Edge score: 1.0
 Number of matched edges: 2
 Sum of edge match values: 2.0
 Number of disconnected matches: 0

Mappings

TG Node	DG Node	Node-Node Score
2	1498	1.0
1	110007	1.0
3	1733	1.0

Attribute Name	Template Node Values	Data Node Values

Template Graph

Attribute Type	TG Value	DG Value
Modifier	IED	IED, trigger
Group		device

Data Graph

Messages

Message Number	Message Content
64	64. 01/27/10 - BCT forces detained a Sunni munitions trafficker after a search of his car netted IED trigger devices. <u>Ahmad Mahmud</u> was placed in custody after his arrest along the Dour'a Expressway, /MGRSCOORD: 38S MB 47959 80868//, in East Dora.
89	89. 02/03/10 - The trigger devices which were netted in the arrest of <u>Dhanun Ahmad Mahmud Ahmadd</u> on 01/27/10 match materials which were found in the truck of arrested ISG affiliate, Abdul Wahied. BCT analysts are attempting to establish their link.
120	120. 02/08/10 - BCT forces interrogating <u>Dhanun Ahmad</u> , arrested on 01/27/10, report his sudden cooperation after being threatened with being turned over to the Shi'a dominated Iraqi National Police (INP). <u>Ahmad</u> admitted to working with high ranking members of a Sunni group operating from Rashid. Ahmad is reluctant to give up any names without some guarantee of his safety. Ahmad claims his business dealings are very tenuous and often cross sectarian lines when profitable. He



Option Year Plans



Situation Assessment: Stochastic Graph Analytics



- **Capability Goal:**
 - Further improve usability of tools and work towards parallelization of additional components
 - Move toward cloud storage and data access (HBase)
- **Research Goals:**
 - Optimize increment batching methodology
 - Allow temporal layering of data graph
 - Enable temporal templates specifying precedence and time range constraints on graph elements
 - Complete uncertainty alignment study to quantify benefit of uncertainty alignment process for graph matching