Challenges in Preserving and Reconstructing Computer-Assisted Decision Processes

Peter Bajcsy, Ph.D. Research Scientist, Image Spatial Data Analysis (ISDA) Group, NCSA Adjunct Assistant Professor, CS and ECE Departments, UIUC

Contributions by Peter Bajcsy, Sang-Chul Lee and Rob Kooper, National Center for Supercomputing Applications (NCSA)



Outline

- Introduction
- Understanding Future Preservation and Reconstruction Requirements
 - Focusing on Information Gathering and Reconstruction
 - Information Gathering Challenges
 - Proposed Approach
 - Simulation Framework for Understanding Preservation and Reconstruction Requirements for Computer Assisted Decision Processes

 Example Tradeoff Studies Using the Simulation Framework

- Basic components: Image Viewer, Event Category Tracker, Event Viewer, Report Generator
- Example Experimental Tradeoff Results
- Discussion



Introduction: Trends

- Two Trends in the Context of Decision Processes
 - Decision processes are moving from paper based to electronic record based (~ computer assisted decision processes)
 - Usability of electronic records depends on rapidly changing information technology
- Implications on Preservation and Reconstruction of Computer Assisted Decision Processes
 - What can we reconstruct from the past?
 - Can we preserve better today?
 - How much is preservation and/or reconstruction going to cost us tomorrow?



Introduction: Challenges

- Top Level Challenges of Preservation & Reconstruction of Computer Assisted Decision Processes
 - Dimensionality
 - Complexity
 - Volatility
 - Cost





- Dimensionality and Complexity Example:
 - Temporal axis (past, today and future processes)
 - Human axis (spectrum of agencies, socio-economics environments, human-computer interactions, ...)
 - who had access/knowledge/resources/connections
 - what, where and when
 - Electronic record axis (spectrum of data formats, document formats, software languages, software platforms, ...)
 - In-silico axis (computational hardware, computers/ peripherals/ networking/...)



Introduction: Challenges

- Volatility Example:
 - Volatility of software/hardware/storage media
 - Updates: Windows operating systems since 2000: Two major new releases, two minor service pack updates, around fifty security patches since SP2
 - **Upgrades:** Microsoft Office Pro for Windows 95/98/ME/2000/XP/2003
 - Media life expectancy: Optical ~5 years, Disk ~ 15 years, Microfiche ~ 100, microfilm ~ 300, newspaper ~ 50, clay tablet ~ 10,000 (life expectancy vs. information density – [P. Conway, 1996])

• Cost Example:

- Cost of software/hardware/storage media
 - Operating System: Windows 3.1/95/98/NT/2000/XP/Vista: Windows 95 = \$209; Windows NT = \$280; Windows XP = \$300; Windows Vista = \$399
 - 128 MB of SDRAM: Year 1999 ~ \$120-> \$40 -> \$200-250 due to Earthquake in Taiwan -> March 2000 ~ \$55->March 2007 ~ \$8.96 (flash card) - <u>www.pricewatch.com</u> (1TB ~\$500)
 - High performance computers: 2006: DARPA awards approximately \$500 million to Cray and IBM; 2007 NSF \$200 million to (NCSA)



Introduction: General Approach

- How can we answer basic questions?
 - What can we reconstruct from the past?
 - Can we preserve better today?
 - How much are preservation and reconstruction going to cost us tomorrow?
- Our Approach: Analysis and Synthesis of Focused Problems
 - Understanding Preservation and Reconstruction of Scientific Simulations Used In Past Computer Assisted Decision Processes
 - Simulation Framework for Understanding Preservation and Reconstruction Requirements of Future Computer Assisted Decision Processes



Simulation Framework for Understanding Preservation and Reconstruction Requirements of Future Computer Assisted Decision Processes



Problem Formulation

- Problem
 - Given a simulation of future computer assisted decision processes,
 - Predict computational requirements for preservation and reconstruction of decision processes

Motivation

- How decision making processes should be documented and archived in the future?
- What computational performance and cost implications would be introduced if decision processes had been documented at the appropriate information granularity?



Information Gathering: Challenges

• Known challenges for building information gathering systems:

- System requirements:
 - Scalable, general and customizable, including trust, security and preservation [Groth et. al., 2004], as well as to providing automatic and systematic provenance gathering [Greenwood et. al., 2003].
- Data sizes:
 - Large amounts and a wide variety of provenance information
- Standards:
 - Paucity of standards, components and techniques for recording provenance [Groth et. al., 2004].
 - Lack of standards in semantic and syntactic descriptions about data leads to difficulties when the data sets are used, stored and retrieved.
 - Short history of standards since the first provenance systems were reported only about a decade ago (Geolineus in 1993 to track ArcInfo GIS operations, GOOSE in 1994, and Geo-Opera in 1997 [Bose & Frew, 2005]).
- Scale:
 - Provenance of a single human (Gordon Bell and his MyLifeBits [Cherry, 2005]) or a community (collaborative environments and experiment management [Bose & Frew, 2005]).
- There is no provenance system that would meet all above requirements, as well as specific requirements defined by multiple application domains.





Proposed Approach

- Divide the problem into three sub-problems
 - Identify a class of computer assisted decision making processes
 - Research information gathering about activities associated with decisions and semi-automated report generation about decision processes
 - Research and evaluate reconstructions of decision activities and authenticity of preserved records
- Develop a simulation framework with three major components
 - Solution to each sub-problem forms one major component
 - Basic archival questions could be studied by prototyping multiple solutions



Simulation Framework Components





Selection of a Class of Computer Assisted Decision Processes

- Objective:
 - Find decision processes requiring common computer functionality
 - Select computer functionality that supports large number of decisions
- Selection:
 - A class of computer assisted decisions based on image inspection
 - Computer functionality to perform image inspections
- Selection Motivation:
 - Image properties
 - Language independent, omnipresent, multi-spectral/multi-dimensional
 - Basic form of important information for many decisions
 - What is of future interest?
 - Was the file viewed?
 - Did the file contain relevant information?
 - Was the image viewed at the appropriate magnification and spatial location?
 - Was image viewed using adequate intensity brightness and color bands?



Simulation Framework



Developed Simulation Framework

- 1. Image reviewer (Decision Process)
 - Contains Image Frame and Image Panel, and includes all image manipulation functionality
- 2. Image event category tracker (Gathering & Summaries & Storage & Reporting)
 - Tracks events in Image reviewer, allows setting preferences for information gathering and storage, lists event activities, summarizes event activities, displays inspected area and displays computational requirements
 - Enables to generate reports with protected authentic information about decision processes, speeds up reporting and provides a foundation for tracking versions of reports
- 3. Image event viewer (Retrieval & Reconstruction & Learning)
 - Retrieves gathered and stored information by Image event category tracker, reconstructs processes with selected information granularity, and displays hierarchy of events and replays image inspection events according to the available information granularity



1. Image Reviewer (Decision Process)



2. Image Event Category Tracker (Gathering & Summary & Storage)

2. Image Event Category Tracker (Semi-Automatic Reporting)

IP2Learn Report - Microsoft Internet Explorer		🎒 IP2Learn Report -	- Microsoft Internet Explo	'er	IP2Learn Report - Microsoft Internet Explorer	
<u>File Edit View Favorites Tools</u>	<u>E</u> ile <u>E</u> dit <u>V</u> iew F	- <u>a</u> vorites <u>T</u> ools <u>H</u> elp		<u>Eile Edit View Favorites Iools H</u> elp	an a	
🕞 Back + 🕥 + 💌 🛃 🦿	🏠 🔎 Search 🦙 Favorites 🕢 🔗 🎍 🗵	🕞 Back 🔹 🌍 🔹	💌 💈 🏠 🔎 Se	arch 👷 Favorites 🚱 🔗 🌭 😿 🔹	🕞 Back 👻 🕥 - 💌 😰 🏠 🔎 Search 🤺 Favorites 🚱 😒 - 🎍	🍃 🗷 • 🛄 🛛 • 🎇 😐 🔘 🦓
Address (C:\PeterB\Projects\NARA)	DOO\Vol018 sub10\test032807\report.html	Address 🙋 C:\PeterB	NProjects\NARA\DOQ\Vol018	sub10\test032807\report.html	Address 🙋 C:\PeterB\Projects\NARA\DOQ\VolD18_sub10\test032807\report.html	💌 🔁 Go 🛛 Links » 🦓 🗸
Google C-	🕇 Go 🖟 🙈 🎉 👻 😚 Bookmarks 🕶 🔊 14 blocked	Google G-	🔻 Go 🚸 🧭) 😵 🕶 🔓 Bookmarks 🕶 🚳 14 blocked	Google 🕞 🗸 🔻 🔽 Go 🖟 🌮 🌑 🏠 🗧 🔂 Bookmarks 🕶 👰 14 b	locked » Osettings - Jenovo
		VI · //·	()	▼ I Search Web ▼ 2 □ ▼ @ ▼ @	1 V - 0 - Search Web - 0 - 6	🕽 - 🕜 🕅 Bookmarks - 🖂 Mail - 🛛 🗎
		ID2Loors Depart	ير ويو الحالي			
IP2Learn Report	Add Tab		Add Ta			
1. Input Image		5. Sequence Ov	erview		SaveSize presents graphically the number of bytes written for each cumulative event and the tot visualization can be selectively adjusted to show only a subset of cumulative events.	al number of bytes written to disk. The 🔝
1251			Table 2:Ten	poral Aggregation of Events: Sequence Overview	2,134,646	<u>ل</u>
ALL STORE		Sequence	ce Event Name Fre	quency Property	2,000,000	
Start St		1	Window Created 5	Multiple		
		2	Change Zoom Factor 1	0.9580000042915344		
		3	Change Visible Region 1	[x=0,y=0,width=500,height=500]		
	the state of the state	4	New Image 1	C:\PeterB\Projects\NARA\DOQ\Vol018_sub		
	10	5	Change Visible Region 1	[x=-1,y=0,width=501,height=500]	1,500,000	
		6	Change Zoom Factor 1	5.0		
		7	Change Visible Region 20	Multiple		
		8	Change Selection 1	[x=15,y=197,width=9,height=1]		
		9	Change Visible Region 18	Multiple		
		10	Window Shown 1	Select Bands	1,000,000	
		11	Change Gray Scale 1	true		
		12	Window Hidden 1	Select Bands		
		13	Change Selection 1	[x=188,y=197,width=15,height=1]		
		14	Change Visible Region 47	Multiple		
		15	Window Shown 1	Select Bands	500,000	
and the second		16	Change Gray Scale 1	false		
The second se	and a state of the second	17	Window Hidden 1	Select Bands		
the second		18	Change Selection 1	[x=392,y=56,width=51,height=62]		
		19	Add Annotation 1	[x=417,y=87,width=51,height=62,label=susp		
10 A		20	Change Zoom Factor 1	2.5	o <u>l</u>	
		21	Change Visible Region 10	Multiple	1,625 100,	000 148,644
Sector States and Sector		22	Window Shown 1	Magnification Dialog	Total Bytes	
Figure 1	l:Input Image(numrows=500, numcols= 500, numbands=3, tj	23	Window Created 3	Multiple	Window Created Change Zoom Factor	
		24	Window Shown 1	Magnifier	Change Visible Region	
2 Annotated Bergiana		25	Change Visible Region 17	Multiple	New Image Change Selection	
2. Funtorated Regions		26	Change Zoom Factor 1	0.9259999990463257	Window Shown	
	0.0	27	Change Auto Zoom 1	true	Change Gray Scale Window Hidden	
	JH1	28	Change Visible Region 1	[x=0,y=0,width=500,height=500]	Add Annotation	
	3023	29	Change Zoom Factor 1	0.9580000042915344	Change Auto Zoom Magnification	
		30	Change Visible Region 2	Multiple	Mouse Clicked	
	Figure 2:suspicious	31	Magnification 3	Multiple	Figure 4: Graphs of the disk space (vertical axis) as a function of ti	ime (horizontal axis) for
1	(Ix-41., y-6., where - 21, not give - 2, table - 5 as pit to as, type - 1	32	Window Hidden 2	Multiple	each event type.	
E Done		🙋 Done			Done	My Computer

3. Image Event Viewer (Retrieval & Reconstruction)

🗜 Event R	econstroctor	[event.zip]						
File View	Tool Help							
Þ	11	Reconstructio	on Method : Cu	stom (Current Se	etting	gs)		~
Time	Event Name	Interpreted	Raw Data	Snapshot		TimeStamp	2006-11-02_13:11:26:835	
2006-11-02	Change Visib	L×=0,y=0,w	rOOABXNyA		~	Name	Change Visible Region	
2006-11-02	New Image	C:\sclee\My	rOOABXNyA	New Image/		- Admo		
2006-11-02	Change Zoo	5.0	rOOABXNyA	Change Zoo	=	interpreted	[x=246,y=1,width=98,height=97]	
2006-11-02	Change Visib	[x=0,y=0,w	rOOABXNyA	Change Visib	-	Raw Event	Available	
2006-11-02	Change Visib	[x=1,y=0,w	rooabxNya	Change Visib		Conschot	Chappe Visible Region/00000016 inc	
2006-11-02	Change Visib	[x=35,y=0,	rOOABXNyA	Change Visib		ShapShot	Change Visible Region/0000016.jpg	
2006-11-02	Change Visib	[x=97,y=0,	rouabxnya	Change Visib				
2006-11-02	Change Visib	Lx=141,y=0	rouabxnya	Change Visib			Martin and a state of the	
2006-11-02	Change Visib	[x=153,y=0	rouabxnya	Change Visib				
2006-11-02	Change Visib	[x=166,y=0	rouabxnya	Change Visib				
2006-11-02	Change Visib	[x=181,y=0	rouabxnya	Change Visib			the second second second second	
006-11-02	Change Visib	[x=191,y=0	rouabxnya	Change Visib			NAME OF TAXABLE AND DESCRIPTION OF TAXABLE	
005-11-02	Change Visib	[x=207,y=0	rouabxinya	Change Visib			A DECK OF THE REAL PROPERTY OF	
006-11-02	Change Visib	[x=226,y=0		Change Visib			AND THE PROPERTY OF A DECISION	
006-11-02	Change Visib	[x=239,y=0	rouadxinya	Change visio	1		SACARDARIA DA MARANA	
2006-11-02	Change Visib	[x=244,y=0	Event V	iewer			Snanshot Viewer	
2006-11-02	Change Visib	[x=246 y=1	LVCIIC V		8		Shupshot viewer	
Daw Event D	ata	LV-51013-1			least.			
Raw Evenic D	ala							10000
Recorded D	ata							- ^
rOOABXN	yACtuY3NhLi	mltMmxlYXJ	uLmNvcmUu	ZG1zcGxheS	5Jb	WFnZVVw2	GFOZUV2ZW5O6xSBt1G6	
WYICAAJ	JAAJpZEwAA	29ianQAEkx	qYXZhL2xh	bmcvT2JqZW	NOC	3hyABVq)	/XZhLnVOaWwuRXZlbnRP	
Ymp 1 Y3 R	MiOlOGG19a	ATAAHhmAAA	AEHNWABJO	YXZhLmF3dC	557	WNOYW5nb	GXDsGoFGsnadAIABEkA	
BmbloUd	odEbiBVdn71	HDOSONBARD		AGENNABINA	110	addddF=	ombbool obpqaninbinn	
Dimiawa	OUEKADAUPZI	IIKOJQADEEK	AAAJIICAAA	AOLAAADIAA	AAD	'YAAAAL-		
-Decoded D/	ata-							-
CHANGE	VISIBLEREG	TON						
java av	t Dectoral							
Java.aw	c.Rectanyi	= DQ heistri	071					
[X=246,	y=1,wiath=9	98, neight=	.97]				Kaw Data Viewer	
String	= java.awt	.Rectangle	:[x=246,y=	l,width=98	, he	ight=97]		
2								Y

Reconstruction from Raw Data

3. Image Event Viewer (Learning) **RDF** Tree Viewer

RDF Triples (Raw)

🖸 event333.rdl - WordPad	
för Edit liven Inset Figmat tjelp	
0	
Chity // Jank. etc.n., use. 4/07/11// provinces/revers. 2 data 24 - 126 - 401 - 402 - 407 Madd Tak (http:// Jank. etc.n. use. 4/07/11/ provinces/revers. 2 data 24 - 126 - 401 - 402 - 407 Madd Tak (http:// Jank. etc.n. use. 4/07/11/ provinces/revers. 2 data 24 - 126 - 401 - 402 - 407 Madd Tak (http:// Jank. etc.n. use. 4/07/11/ provinces/revers. 2 data 24 - 126 - 401 - 402 - 407 Madd Tak (http:// Jank. etc.n. use. 4/07/11/ provinces/revers. 2 data 24 - 126 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 401 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 402 - 40	ststp://ish.ses.uie.ek/UEJ/provessorfhadmes) < fluider & dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/stribes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/stribes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/stribes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/stribes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/stribes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://ish.ses.uie.ek/UEJ/provessorfhadmes/ dttp://is
	And and a second s
rior (1960), press 7 5	N.PI

- D × RDF Viewer [MIevent.rdf] File Help [CDATA[rO0ABXNyACtuY3NhLmlt...]] ٠ 🚊 💼 executeStep event_61f71405-5232-4d17-852b-0e36c7db6220 🚊 💼 hasName New Image' - hasStartTimeStamp in hasInterpreted [CDATA[C:\PeterB\Projects\NARA\DOQ\Vol018_sub10\16tdl080675_sub10.tif]] 🚊 💼 hasRaw [CDATA[rO0ABXNyACtuY3NhLmlt...]] 🚊 💼 hasSnapshot ---- • '[CDATA[New Image/0000000.jpg]]' 💼 executeStep event_d28ccab5-7610-4f09-ae66-c976e862a323

Spatial aggregation of activities based on:

Image Provenance To Learn(IP2Learn) Report

Note: This report is automatically generated by the "Report Generator"

Time duration

1.Frequency Analiysis

NCSA

Example Tradeoff Studies or How to Use the Simulation Framework?

Variables of Tradeoff Simulations (1)

- Information granularity
 - Categories of observable and measurable variables
- Information gathering method
 - Mechanisms how gather information about computer system activities ranging from hardware to user program levels
- Information organization, representation and compression types
 - Multiple 'standard ways' how to organize, represent and compress metadata
- Types of information authentication
 - Several purpose methods: encryption, watermarking or steganography

Variables of Tradeoff Simulations (2)

- Storage formats of gathered information and summary reports
 - Multiple wide spread (standard?) formats of "final" reports
- Information retrieval types
 - Remote or local (distributed or centralized) retrieval, compressed or uncompressed files
- Types of decision process reconstructions
 - Activity levels ranging from human interaction to computer execution
- Reconstruction value metrics
 - Application dependent considerations about reconstruction value

Storage vs. Granularity Tradeoffs

• IP2Learn supports about 26 image inspection operations

- events during image inspection based decision processes

- Three Major Information Granules
 - Interpreted what the programmer encoded as a textual description (interpretation) of an image operation
 - Raw what the computer worked with when an image operation was recorded
 - Snapshots what was rendered on the computer screen at the time of an image operation

Information Granularity

INFORMATION GRANULARITY ABOUT IMAGE OPERATIONS

Example of Information Granules

• Example – Zoom image operation:

- The word "zoom" (interpreted),
- Java event message reporting zoom (raw),
- snapshot of zoomed image (snapshot)

• Actual data:

- Interpreted zoom event:
 - Change Zoom Factor 2.5
- Raw zoom event:
 - Recorded Java event about zoom operation:
 - rO0ABXNyACtuY3NhLmltMmxlYXJuLmNvcmUuZGlzcGxheS5JbW FnZVVwZGF0ZUV2ZW506xSBt1G6WYICAAJJAAJpZEwAA29ianQ AEkxqYXZhL2xhbmcvT2JqZWN0O3hyABVqYXZhLnV0aWwuRXZI bnRPYmplY3RMjQlOGG19qAIAAHhwAAAABHNyABBqYXZhLmxh bmcuRG91YmxlgLPCSilr+wQCAAFEAAV2YWx1ZXhyABBqYXZhL mxhbmcuTnVtYmVyhqyVHQuU4IsCAAB4cEAEAAAAAAA
 - Decoded Java event about zoom operation:
 - CHANGE_ZOOMFACTOR String = 2.5
- Snapshot of zoom event:
 - Change Zoom Factor/0000021.jpg

Storage vs. Granularity Test Case

- Example Image Inspection
 Sequence
 - Defined Task: find a house in a large image located at (machine + directory + file name)
 - Image Inspection Sequence:
 - load a file ->
 - zoom to 50% so that the entire image could fit into a viewable window ->
 - then zoom to 250% so that with the help of a scroll bar the house location can be found ->
 - adjust the gamma intensity value for better visualization.

Storage = function of time, image operation and its granularity

Test Case Conclusions (1)

- 10 out of 26 image operations were used in this test case
- When <u>all information granularities</u> were recorded then Change Visible Region (59.55%) and New Image (29.88%) events dominated the storage requirements of about total 3,600,000 bytes.
- When <u>only 'interpreted' information</u> was recorded then Change Visible Region (70.26%) event required the most storage space and the total space was only about 10,600 bytes.
- When <u>only 'raw' information</u> was recorded then New Image (97.95%) event dominated the total of 1,000,000 bytes of storage.
- When <u>only 'snapshot' information</u> was recorded then Change Visible Region (54.74%), Window Snapshot (27.26%) and Change Zoom Factor (11.30%) were the main contributors to the storage requirements from the total of 547,000 bytes.

Test Case Conclusions (2)

- Estimations of relative storage and execution time requirements:
- The ratios for 'only interpreted' to 'only raw' to 'only snapshot' to 'everything' are
 - (a) approximately 1 : 100 : 55 : 360 for storage requirements and
 - (b) about 1:3.46:16.49:81.42 for execution time requirements.

Storage vs. Information Organization Tradeoffs

Key Pairs

<timestamp>2006-10-26_14:16:16:581</timestamp>

- <interpreted>
- <![CDATA[5.0]]>
- </interpreted>
- <rawevent>
- <![CDATA[

rO0ABXNyACtuY3NhLmltMmxlYXJuLmNvcmUuZGlzcGxheS5Jb WFnZVVwZGF0ZUV2ZW506xSBt1G6WYICAAJJAAJpZEwAA29 ianQAEkxqYXZhL2xhbmcvT2JqZWN0O3hyABVqYXZhLnV0aWw uRXZlbnRPYmplY3RMjQlOGG19qAIAAHhwAAAABHNyABBqYX ZhLmxhbmcuRG91YmxlgLPCSilr+wQCAAFEAAV2Wx1ZXhyAB BqYXZhLmxhbmcuTnVtYmVyhqyVHQuU4IsCAAB4cEAUAAAAA AAA]]>

</rawevent>

- <snapshot file="Change Zoom Factor/00000003.jpg" /> </event>
- <event name="Change Visible Region">
- <timestamp>2006-10-26_14:16:16:800</timestamp>
- <interpreted>
- <![CDATA[[x=0,y=0,width=97,height=97]]]>
- </interpreted>
- <rawevent>
- <![CDATA[

O0ABXNyACtuY3NhLmltMmxIYXJuLmNvcmUuZGIzcGxheS5Jb WFnZVVwZGF0ZUV2ZW506xSBt1G6WYICAAJJAAJpZEwAA29 ianQAEkxqYXZhL2xhbmcvT2JqZWN0O3hyABVqYXZhLnV0aWw uRXZlbnRPYmpIY3RMjQIOGG19qAIAAHhwAAAAEHNyABJqYX ZhLmF3dC5SZWN0YW5nbGXDsGoFGspqdAIABEkABmhlaWdo dEkABXdpZHRoSQABeEkAAXI4cAAAAGEAAABhAAAAAAA AA=]]>

</rawevent>

<snapshot file="Change Visible Region/00000004.jpg" /> </event>

Information Organization

Resource Description Framework (RDF) Triplets

XML Key Pairs: Example

- Key Pairs (tag, value):
 - Example:
 - <event name="New Image">
 - ~(tag1 = event name; value = New Image)
 - <timestamp>2007-03-29_13:28:21:818</timestamp>
 - <interpreted> <![CDATA[C:\PeterB\Projects\NARA\DOQ\Vol018_sub10\16tdl080675_sub 10.tif]]> </interpreted>
 - <rawevent> <![CDATA[rO0ABXNyACtuY3NhLmltMmxlYXJuLmNvcmUuZGlzcGxheS5J bWFnZVVwZGF0ZUV2ZW506xSBt1G6......d254]]
 </rawevent>
 - <snapshot file="New Image/0000000.jpg" />

RDF Triples: Example

- Resource Description Framework (RDF)
 - Representation: Subject Predicate Object format
- Example:
 - Subject = event ID: <u>http://isda.ncsa.uiuc.edu/NARA/provenance#event_61f714</u> 05-5232-4d17-852b-0e36c7db6220
 - Predicate:

<http://isda.ncsa.uiuc.edu/NARA/provenance#hasName>

– Object: <'New Image'>;

Storage Requirements vs. Storage Format Tradeoffs

Storage

- Format:
 - Text: XML
 - Images: JPG or TIFF
- Compression:

Text and Images with or without ZIP

Tradeoff Study Environment	Specifications	
Java Virtual Machine	Version 1.6.0 (build 1.6.0_01-b06)	
Operating System	Microsoft Windows XP Version 2002, Service Pack 2	
Hardware	CPU: Intel Pentium 4 (Northwood) 3.00GHz Memory: 1GBytes	

Test Sequence

Event Name	Frequency	Property	Event Name	Frequency	Property
New Image	1	C:\sclee\My	Change RGB Band	1	[R=-1,G=-1,B=2]
		e\16tdl065660_sub10.tif	Change RGB Band	1	[R=-1,G=-1,B=-1]
Change	1	1.1	Change Gray Scale	1	true
Gamma		1.1	Change RGB Band	1	[R=0,G=1,B=2]
Window	1	Image Info	Change Gray Scale	1	false
Change			Change Zoom Factor	1	1.0
Selection	1	[x=50,y=164,width=50,height =52]	Change Visible Region	1	[x=0,y=0,width=500,heig ht=500]
Magnification	21	Multiple	Change Zoom Factor	1	5.0
Change Selection	1	[x=403,y=42,width=77,height =78]	Change Visible Region	67	Multiple
Mouse Clicked	2	Multiple	Change Selection	1	[x=464,y=453,width=14,h eight=0]
Add Annotation	1	[x=403,y=42,width=77,height =78,label=null,type=2]	Change Visible Region	32	Multiple
Add Annotation	1	[x=50,y=166,width=40,height	Change Zoom Factor	1	1.5479999780654907
	1	=38,label=null,type=2]	Change Auto Zoom	Auto Zoom 1 true	true
Add Annotation	1	[x=98,y=259,width=109,heig ht=163,label=null,type=2]	Change Visible Region	1	[x=0,y=- 1,width=500,height=501]
Add	1	[x=131,y=4,width=51,height	Change Zoom Factor	1	1.5779999494552612
Change RGB	1	[R=0,G=-1,B=-1]	Change Visible Region	1	[x=0,y=0,width=501,heig ht=501]
Change RGB Band	1	[R=-1,G=1,B=-1]	National Center for Supe	rcomputing Appl	

Storage vs. Information Organization Tradeoffs: Test Case

- Information granules include interpreted, raw and snapshots
- Files were not compressed

Storage vs. Compression Test Case

Test Case Conclusions

- Storage Requirements (RDF) > Storage Requirements (KeyPairs)
- Compression ratio (RDF) > Compression ratio (KeyPairs)
- Zipped (KeyPairs and Images) have a small compression ratio
- Uncompressing RDF or KeyPairs takes the same amount of CPU time and it is much less than zipped files
- Storage Requirements (Load New Image) > Storage Requirements (Other events)
- On average, all other events but "Load New Image" took about 2K-5KB while "Load New Image" took about 1MB (200-500 times more)

Authentication of Records

• Records to Authenticate:

- gathered provenance information
- summary reports in HTML (Hypertext Markup Language) or in PDF (Portable Document Format) format
- Methods:
 - Encryption following the Advanced Encryption Standard
 - Watermarking labeling every image with visible text
 - Steganography inserting invisible text or image into preserved images
- Computational requirements of these methods
 - The same test case (image inspection sequence) as before

Computational Requirements vs. Authentication Method Tradeoffs

Test Case Conclusions

- File size does not change with encryption/decryption (as expected)
- Decryption CPU time requirements are slightly higher than the encryption CPU time requirements
- The largest difference between encryption and decryption CPU time is for files with RDF information organization
- Comparison of CPU requirements of authentication methods needs additional considerations

watermarking

Movie: Image Reviewer and Event Tracker

🔛 Event Tracker [event]	🔛 Image Reviewer	
File Option	File Help	
EventViewer SavedViewer RegionTracker SaveSi		1
Date EventName interprete		1. 1. A. 18
Mon Nov 06 14: Window Created Image Info		
Mon Nov 06 14: Window Created Select Zoom	0.0	0
Mon Nov 06 14: Window Created Select Bands		
Mon Nov 06 14: Window Created Select Gamma		
Mon Nov 06 14: Change Visible R [x=0,y=0,wid		
Mon Nov 06 14: New Image C:\PeterB\Pro	i)	
🏄 Start 🗀 3 W. 🗸 🧭 2 I 🗸 🌍 Eud 📝	Ado 🕒 env 💽 2 M 📃 2 J 🛛 🥵 🥵 🖓 🗐 😒 🕲 😨 💬 💂	🧐 🛛 2:33 PM

Movie: Image Event Viewer

Event Reconstroctor [event.zip]	1
File View Tool Help	J
Image: Seconstruction Method : Event Name	zip
Time Event Name Interpreted Raw Data Snapshot TimeStamp 2006-11-06_14:31:25:296	
2006-11-06 Window Cre Image Info	×
2006-11-06 Window Cre Select Zoom	7
2006-11-06 Window Cre Select Bands interpreted Image Info	clipse, exe
2006-11-06 Window Cre Select Gamma Raw Event	
2006-11-06 New Image C/DeterB/Dr rO0ABXNVA New Image/ SpapShot	
2006-11-06 Window Shown Image Info	
2006-11-06 Window Sna Image Info	8
2006-11-06 Window Sna Image Info	(a)
2006-11-06 Window Sna Image Info	a
2006-11-06 Window Sna Image Info	
2006-11-06 Window Sna Image Info	
2006-11-06 Window Closed Image Info	
2006-11-06 Change Zoo 5.0 r(`
Raw Event Data	(7.1
Encoded Data	
Perceded Data	
	dia Player
🏄 Start 🛅 3 W. 🗸 🍠 2 I 🗸 🌍 Eud 📝 Ado 😂 env 🔯 2 M 🖌 Eve 🧕 94‰ 📜 🔍 🕲 🖏 🗒 💬 景 🥹	2:36 PM
National Center for Supercomputing Applicatio	ons NES/

Discussion

- Simulation Framework for Understanding Preservation and Reconstruction Requirements of Future Computer Assisted Decision Processes
 - Image Provenance To Learn (IP2Learn) software with sample data
 - Download: open source & documentation: <u>http://isda.ncsa.uiuc.edu/download/</u>
- Comments? Feedback?
- POC: Peter Bajcsy, pbajcsy@ncsa.uiuc.edu

