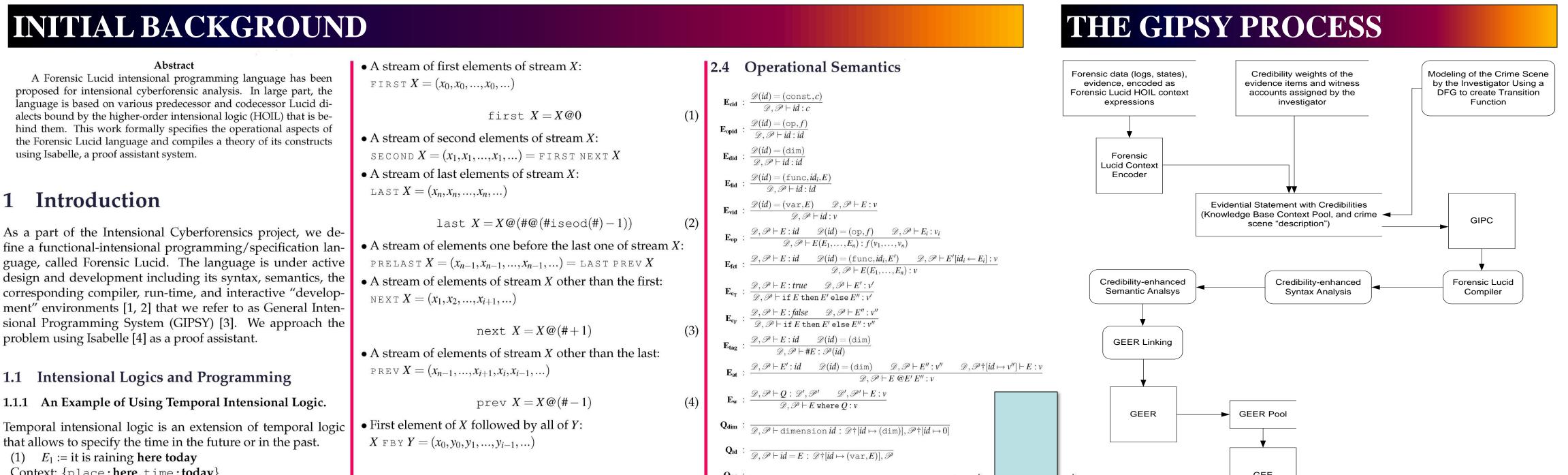
Faculty of ENCS, Concordia University, Montreal, QC, Canada



# The Need to Support of Data Flow Graph Visualization of Forensic Lucid Programs, Forensic Evidence, and their Evaluation by GIPSY

Serguei A. Mokhov, Joey Paquet, Mourad Debbabi

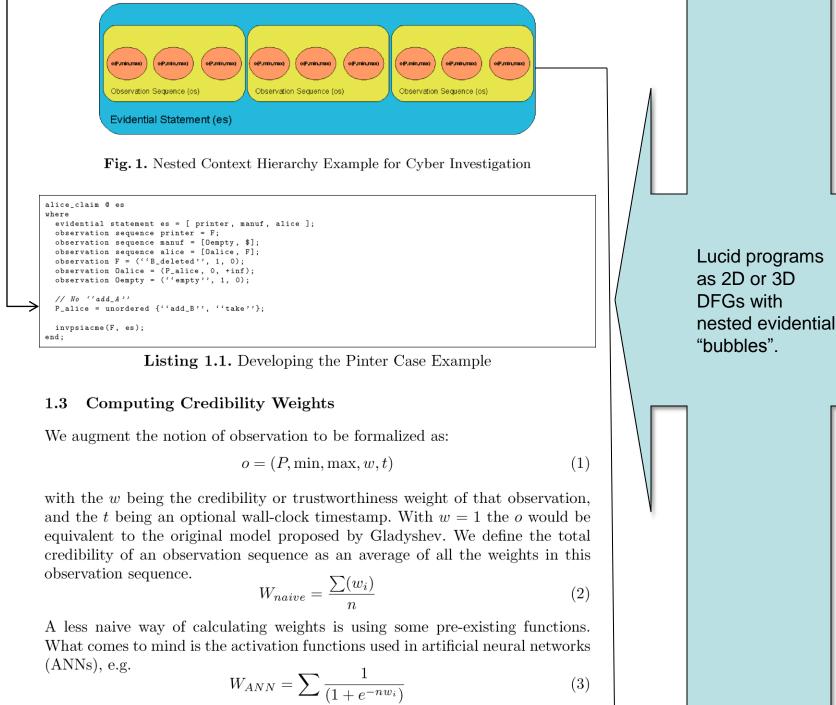


(1) $E_1 := n$ is raining nere today		$\mathcal{D}, \mathcal{J} + \mathfrak{m} = E : \mathcal{D} \setminus [\mathfrak{m} \mapsto (val, E)], \mathcal{J}$	
<pre>Context: {place:here,time:today}</pre>	$\mathbf{Y} \in \{\mathbf{x} \in \mathbf{Y} \mid \mathbf{f} \neq 0 \text{ there } \mathbf{Y} \in \{\mathbf{x} \in \mathbf{Y} \mid \mathbf{f}\}$	$\mathbf{Q_{fid}} \ : \ \overline{\mathscr{D}, \mathscr{P} \vdash id(id_1, \dots, id_n) = E \ : \ \mathscr{D}^{\dagger}[id \mapsto (\texttt{func}, id_i, E)], \mathscr{P}}$	GEE
(2) $E_2 :=$ it was raining here before(today) = yesterday	X fby $Y = \mathbf{if} \# = 0$ then X else $Y@(\# - 1)$ (5)		
(3) $E_3 :=$ it is going to rain <i>at</i> (altitude here + 500 m) <i>af</i> -	= <b>if</b> isbod X <b>then</b> X <b>else</b> prev Y	$\mathbf{Q}\mathbf{Q}:\frac{\mathscr{D},\mathscr{P}\vdash Q:\mathscr{D}',\mathscr{P}' \mathscr{D}',\mathscr{P}\vdash Q':\mathscr{D}'',\mathscr{P}''}{\mathscr{D},\mathscr{P}\vdash QQ':\mathscr{D}'',\mathscr{P}''}$	
ter(today) = tomorrow	• First element of <i>X</i> preceded by all of <i>Y</i> :		
Let's take $E_1$ from (1) above. Then let us fix here to Montreal		Operational	
and assume it is a <i>constant</i> . In the month of February, 2008,	$X$ pby $Y = (y_0, y_1,, y_{i-1},, y_n, x_0)$	$\mathbf{E}_{\mathbf{E}.\mathbf{did}} : \frac{\mathscr{D}(E.id) = (\mathtt{dim})}{\mathscr{D}, \mathscr{P} \vdash E.id: id.id}$	Code Generation
with granularity of day, for every day, we can evaluate $E_1$ to		$\mathscr{D}, \mathscr{P} \vdash E.id: id.id$ encode forensic	
either <i>true</i> or <i>false</i> :	V reper V — if i accord # then V also $V(0)(\# + 1)$ (6)	transition func.	AspectJ-based
Tags: 1 2 3 4 5 6 7 8 9	X pby $Y = $ if iseod # then X else $Y@(#+1)$ (6)	$\mathscr{D} \vdash E : v$ and evaluate.	Evaluation and Evaluation Distributed
Values: F F T T T F F F T	= <b>if</b> iseod <i>Y</i> <b>then</b> <i>X</i> <b>else</b> next <i>Y</i>	$\mathscr{D}, \mathscr{P} \vdash E : v$	Tracing
	• WVR stands for <i>whenever</i> . WVR chooses from its left-hand-		
If one starts varying the here dimension (which could even		type form	
be broken down to $X, Y, Z$ ), one gets a two-dimensional eval-	side operand only values in the current dimension where	dimension (dim)	
uation of $E_1$ :	the right-hand-side evaluates to <i>true</i> .	constant $(const, c)$	
City: / 1 2 3 4 5 6 7 8 9	X wvr $Y=$	operator $(op, f)$	Evaluation of forensic expressions and
Montreal FFTTTFFT	${f if}$ first $Y eq 0$	variable $(var, E)$	probabilities and evidence modeling
Quebec FFFFTTFF	then $X$ fby (next $X$ wvr next $Y$ )	function $(func, id_i, E)$	as nested interconnected objects
Ottawa FTTTTFFF	else (NEXT $X$ WVR NEXT $Y$ )	$\mathscr{P}:\mathbf{Id}\to\mathbf{N}$	
		$\mathscr{P}: \mathbf{Id} \to \mathbf{N}$	
		Each type of identifiers can only be used in the appropriate	
2 Forensic Lucid	$X \text{ wvr } Y = X@T \text{ where} \tag{7}$	situations.	
2 Forensic Luciu	T = U fby $U @ (T+1)$		
	$U = \mathbf{if} V \mathbf{then} \# \mathbf{alse} = \mathbf{ast} + U$	$\mathbf{E}_{\#(\mathbf{cxt})}$ : $\overline{\mathscr{D}, \mathscr{P} \vdash \# : \mathscr{P}}$	<b>Abstract.</b> This work refines the theoretical structure and formal model of the observation tuple with the credibility weight and other factors
The end goal is to define our Forensic Lucid language where	1		for cyberforensic analysis and event reconstruction. The model first pro-
its constructs concisely express cyberforensic evidence, which		$\mathbf{E}_{\text{construction(ext)}} : \frac{\mathscr{D}, \mathscr{P} \vdash E_{d_j} : id_j \qquad \mathscr{D}(id_j) = (\texttt{dim})}{\mathscr{D}, \mathscr{P} \vdash E_{i_j} : v_j \qquad \mathscr{P}' = \mathscr{P}_0 \dagger [id_1 \mapsto v_1] \dagger \ldots \dagger [id_n \mapsto v_n]}{\mathscr{D}, \mathscr{P} \vdash [E_{d_1} : E_{i_1}, E_{d_2} : E_{i_2}, \ldots, E_{d_n} : E_{i_n}] : \mathscr{P}'}$	posed for finite-state automata approach by Gladyshev et al. $[1,2]$ and
can be initial state of a case towards what we have actually ob-	$\bullet$ $\land \circ \land \circ $	$\mathbf{E}_{\text{construction(ext)}} : \frac{\mathcal{D}, \mathscr{P} \vdash [L_{i_j} : V_j]  \mathcal{D} : \mathcal{D} : [V_{i_1} : V_{i_1} : V_{i_1$	later extended and realized in the first iteration of Forensic Lucid, an in-
served as a final state. The implementing system (i.e. GIPSY)	hand-side as a first point in that stream as soon as the right-		tensional forensic case specification language $[3, 4]$ . In the ongoing work, it is being augmented with the Dempster-Shafer theory of mathematical
has to backtrace intermediate results in order to provide the	hand-side evaluates to <i>true</i> .	$\mathbf{E}_{\mathbf{at}(\mathbf{cxt})} \ : \ \frac{\mathscr{D}, \mathscr{P} \vdash E' : \mathscr{P}'  \mathscr{D}, \mathscr{P} \dagger \mathscr{P}' \vdash E : v}{\mathscr{D}, \mathscr{P} \vdash E @ E' : v}$	evidence to include the credibility factors and the like that are lacking
corresponding event reconstruction path, if it exists. The re-	X ASA $Y =$ FIRST $(X$ WVR $Y)$		in the Gladyshev's model. It's practically being implemented within the
sult of the expression in its basic form is either <i>true</i> or <i>false</i> ,			General Intensional Programming System (GIPSY) and the probabilis-
i.e. "guilty" or "not guilty" given the context per explanation			tic model-checking tool PRISM to compile the Forensic Lucid model into the PRISM's syntax and check it with the PRISM tool at run-time.
with the backtrace. There can be multiple backtraces, that cor-	$X \text{ asa } Y = \text{first}(X \text{ wvr} Y) \tag{8}$	3 Conclusion	the FRISM'S Syntax and check it with the FRISM tool at fun-thne.
respond to the explanation of the evidence (or lack thereof).			1 Introduction
	2.3 Forensic Operators	Isabelle is the tool of choice of complete specification of the	1.1 Background
2.1 Proportion	L	mainstream Lucid dialects, and more specifically Forensic Lu-	In one of the first formal approaches about automated cyberforensic case rea-
2.1 Properties	/**	cid, focusing on the correctness of the operational aspects. It is	soning and event reconstruction Gladyshev et al. created a finite-state automata
foo @	* rippena ziech e io each cichicht	a work in progress. The completed work will have a complete	(FSA) model $[1,2]$ to encode the evidence and the stories "told" by witnesses
{		list of the files publicly available and submitted to the AfP [27].	in order to combine them into what they refer to as evidential statement, then
[ final observed event, possible initial observed event ],	*	[	model the FSA of a particular case, and, given the FSA, verify if certain claim
	* @return the resulting combined stream */		agrees with the evidential statement or not and if it does what were possible event sequences that explain that claim. On the other hand, an earlier work
}	combine(s, e, d) =	3.1 Future Work	suggested a mathematical theory of evidence by Dempster, Shafer and others [5,
Listing 1. Intensional Stamphaged Evenession	if iseod s then eod;	The near-future work will consist primarily of the following	6], where factors like credibility, trustworthiness, and the like play a role in
Listing 1: Intensional Storyboard Expression	<pre>else (first s fby.d e) fby.d combine(next s, e, d); fi</pre>	items:	the evaluation of mathematical expressions, which Gladyshev lacked. Thirdly,
			an even earlier work on intensional logics and programming provided a formal
	Listing 2: The <b>combine</b> Operator	• Complete semantics of all the mentioned Lucid dialects and	model that throughout development placed the context as a first-class value in
2.2 Primitive Operators	/**	their formalization with Isabelle.	logical and programming expressions in the system, called Lucid that has pro- duced various I usid dialacts and contact sware systems, such as CIPSV [7, 14]
The collection of the translated exercises denoted in	* Append elements of s2 to element of s1	• Augment the language specification to include the	duced various Lucid dialects and context-aware systems, such as GIPSY [7–14]. Thus, in this work we blend the three together – we augment the Gladyshev's
The collection of the translated operators denoted in	the new pool of the terreteries.	Depmster-Shafer theory [30, 31] of evidence to allow	formalization with the credibility weight and other properties derived from the
monospaced font, while we provide their equivalence to	*/ product(s1, s2, d) =	weights for claims, credibility, belief, and plausibility pa-	mathematical theory of evidence and we encode it as a context in the Forensic
the original Lucid operators, denoted as SMALL CAPS.	if iseod s2 then eod;	rameters.	Lucid language, a Lucid derivative for forensic case management, evaluation, and
The primitive operators are founding blocks to construct	<pre>else combine(s1, first s2) fby.d product(s1, next s2);</pre>	• Prove semantic rules involving intensional data warehouse.	event reconstruction. What follows are the details of our solution along with the
more complex case-specific functions that represent a particu-		• Implementation of the Forensic Lucid compiler, run-time	details of the related work. We intend to translate a Forensic Lucid specification
lar investigation case as well as more complex so-called <i>foren</i> -	Listing 3: The <b>product</b> Operator	1	into the PRISM specification, which is a probabilistic automata evaluation and model checking system [15]
sic operators.		and interactive development environments.	model-checking system $[15]$ .

# **DEVELOPMENTS TO EHANCE FORENSIC LUCID WITH CREDIBILITY AND VISUALIZATION**

#### 1.2 Higher Order Context

HOCs represent essentially nested contexts, e.g. as conceptually shown in Figure 1 modeling evidential statement for forensic specification evaluation. Such a context representation can be modeled as a tree in an OO ontology or a context set. In Forensic Lucid it is expressed following the traditional Lucid syntax with modifications adapted from MARFL [16], e.g. with the main program illustrating the beginning of the Printer Case specification [17] in Listing 1.1.



The witness stories or evidence with higher scores of W have higher credibility. With lower scores therefore less credibility and more tainted evidence.

### The Need to Support of Data Flow Graph Visualization of Forensic Lucid Programs, Forensic Evidence, and their **Evaluation by GIPSY**

Serguei A. Mokhov	Joey Paquet	Mourad Debbabi
Concordia University	Concordia University	Concordia University
Montreal, QC, Canada	Montreal, QC, Canada	Montreal, QC, Canada
mokhov@cse.concordia.ca	paquet@cse.concordia.ca	debbabi@ciise.concordia.ca

#### ABSTRACT

Lucid programs are data-flow programs and can be visually represented as data flow graphs (DFGs) and composed visually. Forensic Lucid, a Lucid dialect, is a language to specify and reason about cyberforensics cases. It includes the encoding of the evidence (representing the context of evaluation) and the crime scene modeling in order to validate claims against the model and perform event reconstruction, potentially within large swaths of digital evidence. To aid investigators to model the scene and evaluate it, instead of typing a Forensic Lucid program, we propose to expand the design and implementation of the Lucid DFG programming onto Forensic Lucid case modeling and specification to enhance the usability of the language and the system.

#### **Categories and Subject Descriptors**

D.1.7 [Programming Techniques]: Visual Programming; D.2.11 [Software Architectures]: Domain-specific architectures; Languages-Forensic Lucid; D.3.2 [Programming Languages]: Language Classifications—Very high-level languages; Multiparadigm languages;; D.3.4 [Programming Languages]: Processors-Compilers; Preprocessors; Run-time environments

#### **General Terms**

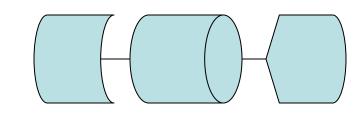
Languages, Design, Security

Keywords

#### Forensic Lucid, DFG, GIPSY, forensic computing

#### 1. OVERVIEW

Cyberforensic analysis has to do with automated or semi-automated processing of, and reasoning about, digital evidence, witness accounts, and other details from cybercrime incidents (involving computers, but not limited to them). Analysis is one of the phases in cybercrime investigation, where the other phases focus on evidence collection, preservation, chain of custody, information extraction that precede the analysis. The phases that follow the analysis are formulation of a report and potential prosecution, typically involving expert witnesses. There are quite a few techniques, tools (hardware and software), and methodologies have been developed for



the mentioned phases of the process. A lot of attention has been paid to the tool development for evidence collection and preservation; a few tools have been developed to aid data "browsing" or the confiscated storage media, log files, memory, and so on. A lot less number of tools have been developed for case analysis of the data (e.g. Sleuthkit), and the existing commercial packages (e.g. Encase or FTK) are very expensive. Even less so there are case management, event modeling, and event reconstruction, especially with a solid formal theoretical base. The first formal approach tc the cybercrime investigation was the finite-state automata (FSA) approach by Gladyshev et. al [5, 4]. Their approach, however. is unduly complex to use and to understand for non-theoreticalcomputer science or equivalent minded investigators. The aim of Forensic Lucid is to alleviate those difficulties, be sound and complete, expressive and usable, and provide even further usability improvements with the GUI to do data-flow graph-based (DFG) programming that allows translation between DFGs and the Forensic Lucid code for compilation. A previous work implemented a simpler solution for Indexical Lucid in the General Intensional Programming System (GIPSY) already [3], but requires forensic and imperative extensions. While Forensic Lucid is in the design and implementation phases, its solid base is being established in part with this work. The goal of Forensic Lucid in the cyberforensic analysis is to be able to express in a program form the encoding of the evidence, witness stories, and evidential statements, that can be tested against claims to see if there is a possible sequence or multiple sequences of events that explain a given story. As with the FSA, it is designed to aid investigators to avoid ad-hoc conclusions and have them look at the possible explanations the Forensic Lucic program "execution" would yield and refine the investigation, as was shown in the works [5, 4] investigators failed to analyze all the stories and their plausibility before drawing conclusions.

## **Related Work**

There as a number of items and proposals in graph based visualization and the corresponding languages. In GIPSY, our own work in the area includes the theoretical foundation and initial practical implementation of the DFGs [17, 3]. Additionally, a part of the proposed related work on visualization and control of communication patterns and load balancing idea was to have a "3D editor" within RIPE's DemandMonitor that will render in 3D space the current communication patterns of a GIPSY program in execution or replay it back and allow the user visually to redistribute demands if they go off balance between workers. A kind of virtual 3D remote control with a mini expert system, an input from which car be used to teach the planning, caching, and load-balancing algorithms to perform efficiently next time a similar GIPSY application is run [7, 10]. Related work by other on visualization of load balancing, configuration, formal systems for diagrammatic modeling

and visual languages and the corresponding graph systems are presented in [19, 18, 1, 2, 6, 16]. They all define some key concepts that are relevant to our visualization mechanisms within GIPSY and its corresponding General Manager Tier (GMT). We propose to build upon those works to represent the nested evidence, crime scene as a 3D DFG, and the reconstructed events flow upon evaluation. Such a feature is projected in the near future to support the previous work of the authors on intensional forensic computing, evidence modeling and encoding, and Forensic Lucid [11, 13, 12, 14, 15] and MARFL [9, 8] (where the intensional hybrid programming languages being realized within GIPSY platform to investigate the language properties and test the run-time). For that related work an example of a 2D DFG corresponding to a simple Lucid program is in Figure 1. In Figure 2 is the conceptual hierarchical nesting of the evidential statement es context elements, such as observation sequences os, their individual observations o, and the properties being observed (P, min, max, w, t) details of which are discussed in the referenced related works. These 2D conceptual visualizations are proposed to be renderable in 3D via an interactive interface to allow modeling complex crime scenes and multidimensional evidence on demand.

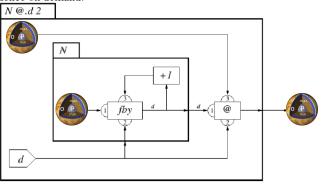
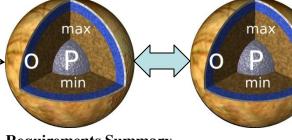


Figure 1: Example of a 2D Data Flow Graph-based Program



#### **Requirements Summary**

Some immediate requires to realize the envisioned DFG visualization of Forensic Lucid programs and their evaluation:

- Visualization of the hierarchical evidential statements.
- Placement of hybrid intensional-imperative nodes into the DFGs. Previous works did not deal with the way on how to augment the DFGAnalyzer and DFGGenerator to support hybrid GIPSY programs. This can be addressed by adding an unexpandable imperative DFG node to the graph. To make it more useful, i.e. expandable and so it's possible to generate the GIPSY code off it or reverse it.
- Java wrapper for the DFG Editor of Yimin Ding [3].
- · Leveraging visualization and control of communication patterns and load balancing for the task in Euclidean space.

#### 2. **REFERENCES**

[1] G. Allwein and J. Barwise, editors. Logical reasoning with diagrams. Oxford University Press, 1996.

- [2] R. Bardohl, M. Minas, G. Taentzer, and A. Schürr. Application of graph transformation to visual languages. pages 105-180, 1999.
- [3] Y. M. Ding. Bi-directional translation between data-flow graphs and Lucid programs in the GIPSY environment. Master's thesis, Department of Computer Science and Software Engineering, Concordia University, Montreal, Canada, 2004
- [4] P. Gladyshev. Finite state machine analysis of a blackmail investigation. Int. J. of Digital Evidence, 4(1), 2005. [5] P. Gladyshev and A. Patel. Finite state machine approach to
- digital event reconstruction. Digit. Investig. J., 2(1), 2004. [6] N. G. Miller. A Diagrammatic Formal System for Euclidean Geometry. PhD thesis, Cornell University, U.S.A, 2001.
- [7] S. A. Mokhov. Towards hybrid intensional programming with JLucid, Objective Lucid, and General Imperative Compiler Framework in the GIPSY. Master's thesis, Department of Computer Science and Software Engineering,
- Concordia University, Montreal, Canada, 2005. [8] S. A. Mokhov. Encoding forensic multimedia evidence from MARF applications as Forensic Lucid expressions. In CISSE'08, pages 413-416. Springer, Dec. 2008.
- [9] S. A. Mokhov. Towards syntax and semantics of hierarchical contexts in multimedia processing applications using MARFL. In COMPSAC, pages 1288-1294. IEEE CS, 2008. [10] S. A. Mokhov. Hybrid Intensional Computing in GIPSY: JLucid, Objective Lucid and GICF. Lambert Academic Publishing, Mar. 2010. ISBN 978-3-8383-1198-2.
- [11] S. A. Mokhov, J. Paquet, and M. Debbabi. Formally specifying operational semantics and language constructs of
- Forensic Lucid. In IMF'08, pages 197-216. GI, 2008. [12] S. A. Mokhov, J. Paquet, and M. Debbabi. Reasoning about a simulated printer case investigation with Forensic Lucid. In
- HSC'09. SCS, 2009. To appear. [13] S. A. Mokhov, J. Paquet, and M. Debbabi. Towards automated deduction in blackmail case analysis with
- Forensic Lucid. In HSC'09. SCS, 2009. To appear. [14] S. A. Mokhov, J. Paquet, and M. Debbabi. Towards automatic deduction and event reconstruction using Forensic Lucid and probabilities to encode the IDS evidence. In
- RAID'10, LNCS 6307, pages 508-509. Springer, 2010. [15] S. A. Mokhov and E. Vassev. Self-forensics through case studies of small to medium software systems. In IMF'09,
- pages 128-141. IEEE CS, 2009. [16] OpenESB Contributors. BPEL service engine. [online],
- 2009. https: //open-esb.dev.java.net/BPELSE.html.
- [17] J. Paquet. Scientific Intensional Programming. PhD thesis, Department of Computer Science, Laval University, Sainte-Foy, Canada, 1999. [18] P. C. Vinh and J. P. Bowen. On the visual representation of
- configuration in reconfigurable computing. Electron. Notes Theor. Comput. Sci., 109:3-15, 2004. [19] C. Zheng and J. R. Heath. Simulation and visualization of
- resource allocation, control, and load balancing procedures for a multiprocessor architecture. In IASTED'06, pages 382–387. ACTA Press, 2006.

VizSec 2010, Ottawa, ON, Canada, September 14, 2010