Towards Automatic Deduction and Event Reconstruction Using Forensic Lucid and Probabilities to Encode the IDS Evidence

Serguei A. Mokhov, Joey Paquet, and Mourad Debbabi

Concordia University, Montréal, Québec, Canada,
{mokhov,paquet,debbabi}@encs.concordia.ca

Introduction. We apply the theoretical framework and formal model of the observation tuple with the credibility weight for forensic analysis of the IDS data and the corresponding event reconstruction. Forensic Lucid – a forensic case modeling and specification language is used for the task. In the ongoing theoretical and practical work, Forensic Lucid is augmented with the Dempster-Shafer theory of mathematical evidence to include the credibility factors of the evidential IDS observations. Forensic Lucid's toolset is practically being implemented within the General Intensional Programming System (GIPSY) and the probabilistic model-checking tool PRISM as a backend to compile the Forensic Lucid model into the PRISM's code and model-check it. This work may also help with further generalization of the testing methodology of IDSs [10].

Overview. Encoding and modeling large volumes of network and other data related to intrusion detection with Forensic Lucid for the purpose of event correlation and reconstruction along with trustworthiness factors (e.g. the likelihood of logs being altered by an intruder) in a common specification of the evidential statement context and a digital crime scene is an important step in the incident analysis and response. One goal is to able to collect the intrusion-related evidence as the Forensic Lucid's evidential statement from diverse sources like Snort, netflows, pcap's data, etc. to do the follow up investigation and event reconstruction. Another goal is to either be interactive with an investigator present, or fully automated in an autonomous IDS with self-forensics [9].

Background. In the first formal approach about automated cyberforensic case reasoning and event reconstruction, Gladyshev et al. created a finite-state automata (FSA) model [3] to encode the evidence and witness accounts of an incident in order to combine them into an *evidential statement*. Then, they modeled the FSA of a particular case, and, verified if certain claim agrees with the evidential statement, and if it does, list possible event sequences that explain that claim [3]. This was followed by the formal log analysis approach by Arasteh et al [1]. Another earlier work suggested a mathematical theory of evidence by Dempster, Shafer and others [4,12], where factors like credibility play a role in the evaluation, which Gladyshev lacked. Thirdly, another earlier work on intensional logics and programming provided a formal model that throughout its evolution placed the context as a first-class value in language expressions in the system, called Lucid that has produced various Lucid dialects and context-aware systems, such as GIPSY [2,13,11]. Thus, we blended the three together – we augmented the Gladyshev's formalization with the credibility weights and we encode the IDS evidence as a higher-order context (HOC) in the Forensic Lucid language. We then translate a Forensic Lucid specification into the PRISM specification, which is a probabilistic automata evaluation and model-checking system and building a PoC expert system bound to it in CLIPS. Some own work done includes [7,5,8,9].

Computing credibility weights. The notion of an observation is formalized in Equation 1 where w is the credibility weight of that observation, and t is an optional wall-clock timestamp. With $w = 1$ the o would be equivalent to the original model proposed by Gladyshev. We then define the total credibility of an observation sequence as an average of all the weights in this observation sequence. The IDS evidence with higher scores of W have higher credibility.

$$o = (P, \min, \max, w, t) \tag{1}$$

$$W_{naive} = \frac{\sum(w_i)}{n} \tag{2}$$

Higher-order context. HOCs represent nested contexts, e.g. as shown in Equation 3 by modeling the evidential statement es containing observation sequences os containing observations o for forensic specification evaluation. In Forensic Lucid it is expressed following the traditional Lucid syntax with modifications adapted from MARFL [6].

$$\boxed{es \; \boxed{os_1 \; \boxed{o_1 \; \boxed{(P, min, max, w, t)}} \; \boxed{o_2} \; \boxed{o_3} \; \boxed{\ldots}} \; \boxed{os_2} \; \boxed{os_3} \; \boxed{\ldots}} \tag{3}$$

References

1. Arasteh, A.R., Debbabi, M., Sakha, A., Saleh, M.: Analyzing multiple logs for forensic evidence. Digital Investigation Journal 4(1), 82–91 (Sep 2007)
2. Ashcroft, E.A., Faustini, A., Jagannathan, R., Wadge, W.W.: Multidimensional, Declarative Programming. Oxford University Press, London (1995)
3. Gladyshev, P., Patel, A.: Finite state machine approach to digital event reconstruction. Digital Investigation Journal 2(1) (2004)
4. Haenni, R., Kohlas, J., Lehmann, N.: Probabilistic argumentation systems. Tech. rep., Institute of Informatics, University of Fribourg, Fribourg, Switzerland (Oct 1999)
5. Mokhov, S.A.: Encoding forensic multimedia evidence from MARF applications as Forensic Lucid expressions. In: CISSE'08. pp. 413–416. Springer (Dec 2008)
6. Mokhov, S.A.: Towards syntax and semantics of hierarchical contexts in multimedia processing applications using MARFL. In: COMPSAC. pp. 1288–1294. IEEE CS (2008)
7. Mokhov, S.A., Paquet, J., Debbabi, M.: Formally specifying operational semantics and language constructs of Forensic Lucid. In: IMF'08. pp. 197–216. GI (Sep 2008)
8. Mokhov, S.A., Paquet, J., Debbabi, M.: Reasoning about a simulated printer case investigation with Forensic Lucid. In: HSC'09. SCS (Oct 2009), to appear
9. Mokhov, S.A., Vassev, E.: Self-forensics through case studies of small to medium software systems. In: Proceedings of IMF'09. pp. 128–141. IEEE Computer Society (Sep 2009)
10. Otrok, H., Paquet, J., Debbabi, M., Bhattacharya, P.: Testing intrusion detection systems in MANET: A comprehensive study. In: CNSR'07. pp. 364–371. IEEE CS (2007)
11. Paquet, J., Mokhov, S.A., Tong, X.: Design and implementation of context calculus in the GIPSY environment. In: COMPSAC 2008. pp. 1278–1283. IEEE CS (Jul 2008)
12. Shafer, G.: The Mathematical Theory of Evidence. Princeton University Press (1976)
13. Wan, K.: Lucx: Lucid Enriched with Context. Ph.D. thesis, Department of Computer Science and Software Engineering, Concordia University, Montreal, Canada (2006)