C6 On Generic Digital Forensic Readiness

Martin S. Olivier, PhD, University of Pretoria, Pretoria, GP 0028, SOUTH AFRICA*

Learning Overview: The goal of this presentation is to move the discourse on digital forensic readiness from how such information may be collected to a deeper discussion of the cost/benefit trade-offs required (where cost also refers to the privacy of the innocent).

Impact on the Forensic Science Community: This presentation will impact the forensic science community by presenting a simple, generic, digital forensic readiness model that allows researchers to propose specific readiness models more concisely. More importantly, the concise description facilitates comparison. In particular, it encourages deeper reflection on the nature, utility, and impact of proactive traces collected.

In digital forensics the phrase "forensic readiness" refers to information that needs to be collected during day-to-day operations of IT systems, such that the evidence required to examine a specific case at some stage will be available and known to be reliable. This presentation provides a generic model of the notion. The intention is to move the discourse from *how* such information may be collected to a deeper discussion of the cost/benefit trade-offs required (where cost also refers to the privacy of the innocent).

Had the phrase "forensic readiness" occurred in other forensic science disciplines, it would probably have referred to the availability of data and samples to facilitate a laboratory's (or analyst's) ability to examine a variety of cases. Examples that come to mind are databases of fingerprints, fiber characteristics, and chemical compositions of drugs, to name but a few. The hashes of known software maintained as part of the United States National Software Reference Library is arguably the best-known example of such preparation in the digital forensics discipline.[1]

Note that such "readiness" is not entirely foreign to forensic science. Cockpit voice and data recorders are of immense value when the causes of aviation accidents are examined (and they are present solely for such investigations). In many contexts a (manual or digital) log of activities is maintained that, again, is very useful during an investigation that involves those activities. However, a more abstract (and more formal) description of typical forensic readiness not only serves to better distinguish such work from the examples mentioned above, but also enables one to present a generic forensic readiness model.

In papers on forensic readiness, such readiness is typically engineered for some system S. The readiness often prepares for some irregularity (such as specific crimes or contraventions of corporate policy). Let i be some irregular activity (such as spoofing of an email or some specific form of fraud). The set of activities A_i that would be sufficient to perform i is then determined. To be ready to examine whether i occurred, it is posited that each activity in A_i should leave a trace. Such a trace is recorded in a logging facility, with the nature of the log entry dependent on the information required to prove i. Let, for any set of traces, T, the proposition $\rho(T,i)$ denote that T is sufficient to prove i. If $\rho(T,i)$, then it follows that $\rho(T,\neg i)$.

This study contends that a simple model based on this notation simplifies the description of a forensic readiness model. The papers explore this for several published readiness models.

Based on the simple model, important questions come to the fore, such as whether the space required for storing traces is warranted by the prevalence (or impact) of any given inappropriate action i. Questions about the size of the generated traces naturally raise the question of whether the size for any proposed mechanism is minimal. It may be possible to offset the costs of being ready for i if i-readiness also implies j-readiness for an irregular activity j; expressed formally, that is when $\rho(T,i) \Rightarrow \rho(T,j)$.

A major concern about readiness models is the fact that they collect "evidence" about innocent people even before an irregular activity is performed. It is possible to consider privacy metrics for a set of traces T. If two equivalent readiness models lead to the collection of traces T and T^0, respectively, then the one with the better privacy score is obviously the better choice. However, the nature of such privacy metrics requires further research.

In summary, this presentation presents a simple generic digital forensic readiness model that allows researchers to propose specific readiness models more concisely. More importantly, the concise description facilitates comparison. In particular, does it encourage deeper reflection on the nature, utility, and impact of proactive traces collected?

Reference(s):

1. Rowe, N.C. 2012. Testing the National Software Reference Library. *Digital Investigation* 9:S131–S138. *The Proceedings of the Twelfth Annual DFRWS Conference.*

Digital Forensic Readiness, Logging, Cost of Extensive Logging