Towards a Relationship-Centric Knowledge Representation Framework for Situational Awareness in Computer Network Defence

L. I. Lumb, H. Hendrawan and A. K. Ho
{ianlumb, hendra, aho}@yorku.ca
University Information Technology
York University, Toronto, Ontario, Canada

Preparation, identification, containment, eradication, recovery and lessons learned are the six steps that characterize traditional incident response. Owing to their potential for impact, however, traditional-incident-response efforts ran the risk of being hyper-focused on computer worms in the past (e.g., the Blaster worm of 2003) and phishing in the present ... while potentially ignoring almost everything else that may be of concern. Of course rather than constantly assuming the reactive posture of traditional incident response, proactive approaches that emphasize prevention would appear to have merit. In the higher-education context, however, preventative measures become rapidly impractical owing to:

- The static-to-declining funds available for the significant one-time-only capital outlays required to acquire special-purpose solutions (e.g., firewalls, intrusion-prevention systems, anti-virus solutions, etc.), and perhaps even more challenging, ongoing operation of the same; plus
- Porous and fuzzy trust boundaries - whose porosity and fuzziness is increasingly exacerbated by freedoms and entitlements that span intellectual to personal (e.g., wired/wireless, fixed/mobile device choice) grounds.

These latter challenges with prevention, in tandem with those formerly identified with respect to traditional incident response, have had the combined effect of shifting emphasis towards detection and response in the higher-education sector. In order to be an effective, reactive strategy for Computer Network Defence (CND), however, approaches based on detection and response need to be highly automated and highly efficient. Thus the promise of a highly automated and highly efficient detection and response scheme is the ability to largely address BotNets, spyware, leveraged/low and slow hacking, plus large-scale virus/worm infestations, while information-security staff remain largely available for situations that really require their involvement.

In concert with preventative measures then, automated and efficient detection and response holds the potential for addressing the CND requirements of the environments that typically exist in the higher-education sector. However, deriving measures of automation and efficiency in this CND context proves challenging on a number of fronts. In the current effort, emphasis is placed two of these challenges. Firstly, for detection to be effective, data must be gathered from all available sources - e.g., logs and honeypots. From the generic and ubiquitous syslog, to device/appliance-specific (e.g., flow, intrusion detection/prevention, firewall, captive portal, etc.) logs, a plethora of sources make data available (please see the left-hand side of Figure 1). Unfortunately, however, even in cases where standards uptake is in evidence (e.g., RFC 3164

for the syslog protocol), payloads of interest in a CND context need to be deftly extracted by detection agents.

Assuming that ‘signals’ of CND interest can be extracted from the ‘noise’ of the tsunami of data, the second challenge that must be addressed is that of attempting to make sense of the same. Broadly speaking, the objective of event correlation is to rapidly identify the actors and their actions in a specific incident context - e.g., the process of mapping events to a specific Internet Protocol (IP) or device (i.e., MAC) address.

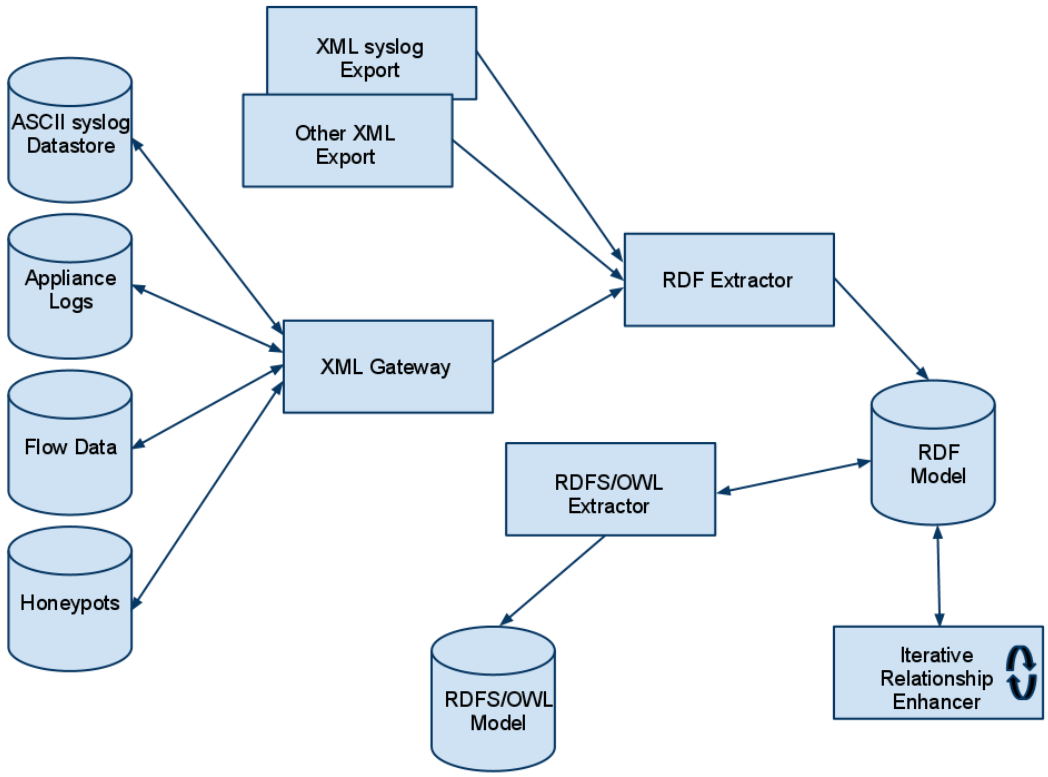

Figure 1. Schematic of the knowledge-representation framework that systematically enhances the expressibility of situational CND data into CND information and ultimately CND knowledge (after Lumb et al., 2010).

In focusing attention here on detection and correlation, it is necessary to ignore aspects of a broader solution - a solution that would likely include the ability to perform various communicative (e.g., notification, logging, ticketing, etc.) and corrective (e.g., network disconnection of an identified device) actions. Specifically, the purpose of the current investigation is to apply a Knowledge-Representation Framework (KRF) to the challenges extant in situational detection and correlation. With respect to detection, the intention is to make use of this KRF to enhance the *expressibility* of the data that is being gathered. In other words, through an iterative process, the purpose is to make explicit relationships that can be derived by direct

parsing, for example, of syslog data via a detection-enhanced XML (eXtensible Markup Language) gateway (please see Figure 1).[1] In this KRF, relationships are extracted (please see "RDF Extractor" in Figure 1) from the XML-based representation and then cast as subject-predicate-object triples in the Resource Description Framework (RDF). The set of all RDF triples comprises an RDF model (please see Figure 1) of the CND situation under consideration.

In a more-expressible and accessible relationship-centric format, the resulting semantically enriched *information*, can be more effectively manipulated. An aspect of this manipulation, is that of performing correlation. Unlike other efforts, the starting point for correlation in the current effort is that of explicitly expressed relationships. Because they have been ultimately derived from original data sources, the relationships so extracted have the potential to be quite objective. Using the KRF, these RDF-expressed relationships undergo a further enhancements expressibility that result in extracted schemas based on RDF (RDFS) and ultimately informal ontologies based on the Web Ontology Language (OWL) (please see "RDFS/OWL Model" in Figure 1). This data-driven approach to the construction of an informal ontology for a CND situation marks a significant point of contrast for the present work from that that already exists in the literature - i.e., approaches based on the use of formal ontologies. That stated however, and as has been demonstrated elsewhere, the incorporation of existing data formats (e.g., ASCII, binary, XML, etc.), and even integration of ontologies themselves, has already been demonstrated as a strength of the KRF (e.g., Lumb et al., 2010). The real-time, on-demand nature of the current approach based on a KRF shares clear affinities with the need for automated and efficient detection and response in a CND context. In addition to applying the KRF that has already demonstrated utility in various scientific (Lumb et al., 2009) and IP network management (Lumb et al., 2010) applications, the current effort demands that the KRF be extended in various ways - broadly speaking, to be more receptive to the real-time, on-demand needs typical of situational CND.

Situational awareness in the context of CND provides both a rich opportunity to apply and present requirements for the KRF. Such a contribution should be included in a book dedicated to presenting the state-of-the-art in this subject area.

References

L. I. Lumb, J. R. Freemantle, J. I. Lederman, and K. D. Aldridge. Annotation modeling with formal ontologies: Implications for informal ontologies. *Comp. & Geosci.,* **35**, 855–861, 2009.

L. I. Lumb, R. Gorsht & D. Zeng, Knowledge Maps for Campus IP Networks: From Relational Databases to Relationship-Centric Semantic Models, K. A. Haffner (ed.), **Semantic Web: Standards, Tools and Ontologies**, Nova Science Publishers, Inc., 67-125, 2010.

[1]Also illustrated schematically in Figure 1 are systems (e.g., both Cisco IOS and Juniper JUNOS platform switches and routers plus Oracle audit trails) that export syslog data natively in XML.