Designing IDLE: The Intrusion Data Library Enterprise*

Ulf Lindqvist[1,3] Douglas Moran[2] Phillip A. Porras[3] Mabry Tyson[2]

[1]Dept. of Computer Engineering
Chalmers University of Technology
Göteborg, Sweden
e-mail: ulfl@ce.chalmers.se

[2]Artificial Intelligence Center
SRI International
Menlo Park, California
{moran, tyson}@ai.sri.com

[3]Computer Science Laboratory
SRI International
Menlo Park, California
{ulf, porras}@csl.sri.com

High-quality, timely information on intrusions is crucial in the development, testing, tuning, and updating of intrusion detection systems (IDSs) and intrusion recovery systems. We present the Intrusion Data Library Enterprise (IDLE), a design and initial compilation of an extensible library of intrusion data that is efficiently parseable in both human-readable and platform-independent machine-readable forms. We are currently in the early stages of design and are building example entries. The IDLE library will be made available as a resource specifically for the intrusion detection community. IDLE will provide IDS developers and users with accurate field data for testing and tuning, and as new intrusion types are discovered, it will enable tools to automatically update rulesets and parameters.

Our experience from detailed intrusion analysis indicates that a vast amount of information is needed about the particular intrusion sample to make correct statements about an intrusion in terms of prerequisites, impact, traces, difficulty, remedies etc. Also, because systems are continually patched to block known intrusions, it can be difficult to recreate a vulnerable system configuration for each intrusion sample when detailed vulnerability information is missing. A major problem facing IDS developers is that the intrusion databases utilized provide little leverage for automating extensions to their systems. Legitimate concerns about distributing information on intrusion schemes has hampered the growth of these databases. Ironically, this has led to the current situation where security professionals find their best sources for intrusion data on underground Internet sites, even though the information published on such sites is often incomplete and of varying reliability.

In IDLE, we attempt to create a standard format that will facilitate rapid distribution of information among IDS developers and related groups in order to achieve "critical mass" in the coverage. However, the breadth of such exchange and the access controls are outside the scope of this work. The emphasis of our design is to include information that ensures automated repeatability, detection, and diagnosis of each intrusion sample. What makes IDLE different from current intrusion databases is that it is designed to serve the IDS community by coordinating detailed information on vulnerable configurations and exploit instructions with documented observable dynamic and static traces (signs) of the intrusion. The IDLE trace information is structured in a form that will support an IDS downloading a new description and extracting the information needed to automatically generate new rules (signatures, parameters, etc.) to identify the new intrusion.

Support for partial information is a core part of the design of IDLE. Information about an intrusion will typically be initially incomplete, and different groups may tend to populate only chosen subsets of a record. Incremental population is especially important in the observables, because different developers monitor system activity from different perspectives: network traffic, audit logs, application logs, filesystem traces, etc. IDLE must also be easily extensible to support the aspects relevant to new and different target platforms, tools, and IDSs.

We have chosen to use the Extensible Markup Language (XML), successor to HTML, for the intrusion database. There is an intense ongoing development of tools for authoring, displaying, browsing and handling XML documents, and we expect substantial leverage from this activity. XML provides a number of features such as platform-independence, naturally hierarchical structure, customizable field display filtering, the possibility to mix human-readable free-text fields and machine-readable fields in the same record, and easy addition of new types of fields. Those features enable IDLE to evolve both in terms of the content of individual records and of the structure of the library. This makes us confident that for IDLE, XML is a far better choice than an existing proprietary database format. □

*Presented at RAID '98 (First International Workshop on the Recent Advances in Intrusion Detection), September 14–16, 1998, Louvain-la-Neuve, Belgium.