



End-to-End Diagnostic Discovery Toolkit

An architecture and beta toolkit supporting integrated analysis and diagnosis of distributed, layered, and interdependent components and systems.

Imagine sitting down at your computer, ready to submit the grant proposal you've been working on all weekend with hours to spare before the noon deadline. You bring up the funding agency's web site in your browser, add the appropriate information, and click the submit button. There's no response. You click again, and still nothing. Did it get submitted?

Your problem could be caused by perhaps a hundred different circumstances: maybe the web site is at fault, maybe there's an intervening firewall, maybe the link is down, or your document has a virus. How can you quickly pinpoint the most effective next step as you try to fix (or work around) the difficulty?

The Diagnostic Challenge

In this scenario, the systems and network are probably being monitored and collecting log information, some of which might even pinpoint the source of your problem and indicate an effective response. The challenge is that the log information is never written in the language of the problem. There's no log that says "Doug doesn't know if he submitted his proposal," so it's incredibly difficult to map the question (what went wrong) to the solution (here's what happened and what you need to do next), even in the simplest of scenarios.

So why is this such a hard problem?

- Though there is a lot of log and diagnostic data being collected, it is seldom managed to enable access to those who need it, when they need it.
- Campuses typically collect a lot of log and diagnostic data, and it's hard to prioritize and focus on the bits that really matter.
- Often the data that is collected doesn't quite fit the question you're trying to ask.

The EDDY Solution

The End-to-End Diagnostics Discovery (EDDY) Architecture provides a framework to begin thinking about the problem. It does this by defining a way to effectively bring together information about activities and anomalies in an infrastructure to enable integrated analysis, research, and audit. Using EDDY-based tools, a system or network manager could more easily discover and diagnose problems as they occur, allowing independent processes to assist in their prediction, management, maintenance, and circumvention. This release of the EDDY Toolkit enables early experimentation with semantic possibility and engineering value in applying this approach.

By itself, EDDY doesn't solve any of the three problems offered in the section above, but an EDDY-enabled diagnostic infrastructure in any environment would provide a significant new opportunity to begin to address them.

What's Included in the Current Release?

As a beta-release toolkit, the EDDY components are designed to foster engineering and experimentation in data correlation and diagnostic discovery. The components that are available in the 2007 release include:

- Additional Common Event Record (CER) to normalize information from a variety of sensors or disparate logging systems. This enables correlation of events amongst sub-systems or across various layers in the infrastructure; whether occurring on the network, in an application, on a particular computer system, or in the operations center that houses them. In this release, three new types of CERs include generic logging for experimentation, Syslog, and Shibboleth.
- An event-dissemination environment (EDDY Backplane) to distribute events across a diagnostic framework, enabling collection, correlation, and analytics. This allows a campus to track and manage data about whatever infrastructures and services it offers.
- A diagnostic tool development platform that allows a developer to easily bring data to their work with new analytic methods they want to develop.
- A definition of the CER Factory architecture to support template-based definition of EDDY CERs using light-weight description methods.
- An updated software components to achieve greater efficiencies in XML processing with JAXB 2.0.

A campus can experiment with EDDY without impacting its existing log and diagnostic infrastructure, either by leveraging new EDDY capabilities for existing data, or by adding EDDY data to current diagnostic capabilities and tools.

Included in the EDDY toolkit are a few sample agents that developers can use directly or tailor to their own needs, adding a new monitor or analysis tool as appropriate. Additional information about new data formats and related development efforts are available on the website and by contacting the EDDY Team.

The current version of the EDDY Toolkit is intended to spur experimentation and innovation and encourage on-going broad participation in this effort. In conjunction with this fourth release, the EDDY Team is helping to foster coordinated experimentation and leverage learning in diagnostic infrastructures and analytics.

Acknowledgements

Development of the EDDY Toolkit was supported with funding from Carnegie Mellon University and the National Science Foundation (CNS-0433540), with early support from Internet2, the NSF Middleware Initiative (Cooperative Agreement No. OCI-0330626).