


CS486 - Senior Capstone Proposal

Project Title: Zeek Package Repository Website	
Sponsor Information: 	Tim Wojtulewicz, Senior Software Engineer Corelight, Inc. and the Zeek open-source team tim@corelight.com 480-325-8953

Project Background:

From the Zeek documentation:

Zeek is a passive, open-source network traffic analyzer. Many operators use Zeek as a network security monitor (NSM) to support investigations of suspicious or malicious activity. Zeek also supports a wide range of traffic analysis tasks beyond the security domain, including performance measurement and troubleshooting.

Zeek implements a system for collecting data about traffic on a network, performing real-time analysis on the traffic, and outputting a series of log data to disk. These logs enable threat hunters to both detect some forms of problematic behavior as it happens, as well as provides data for later forensic analysis. The default set of analyzers covers a large number of the major network protocols and file formats. It also implements it's own scripting and event language, known as Zeek Script, for extending Zeek's capabilities.

One of the major features of Zeek is a plugin architecture that allows third-party developers to implement new analyzers, scripts, and features in Zeek without needing to modify the core code directly. Along with that, Zeek also provides a system for bundling these plugins and such into standalone, easy-to-install packages that anyone can download and add to their setups. More information about plugins and packages can be found at <https://docs.zeek.org/en/current/devel/plugins.html>. The Zeek Project provides a registry for first- and third-party packages at <https://packages.zeek.org>. Developers can open pull requests against a GitHub repository at <https://github.com/zeek/packages> to create and update package entries. A workflow job runs automatically to update the site with the new information once their request has been merged by a Zeek maintainer.

Project Overview:

With the above background now in mind, we can get to the overview for what we're proposing for the project. In a nutshell, we need a new package site and potentially some upgrades to how

the package system works in general. The existing site is fairly dated and somewhat difficult to maintain, and lacks features that developers are used to from package systems available to other languages.

Core functionality:

- Static files generated automatically from the package metadata. These don't necessarily have to be HTML. For example, the package manager could just generate JSON and the site could generate the HTML from that for display.
- Preserve the look-and-feel from the zeek.org website.
- Pulling package descriptions directory from GitHub, such as from a README.md in the repository. The existing site does this already but we don't properly handle images linked in those descriptions. The could either be downloaded locally or the links could be fixed to be absolute links back to GitHub.
- Adding metadata to zkg packages about what versions of Zeek a package is known to work with. This will require some changes to the package template.
- Better searching. It'd be neat if this could support GitHub-like search tags for searching metadata within the packages.
- Linting of the packages at parse time, allowing us to assign a quality score to a package based on a set of metrics (such as good tags, has a license, has a build command, has tests, etc).

Bigger goals:

- Automated testing of packages when a package is submitted to the packages repository. This is easier to do with script-only packages, but hard with packages requiring compilation especially if they have external dependencies.
- Expansion of the linter. The Zeek project has a tree-sitter library (<https://github.com/zeek/tree-sitter-zeek>) for parsing scripts into easily-digestible metadata. This would allow us to automatically add a lot of metadata to a package, such as what events are processed by it.
- Some way to generate usage-based statistics for packages. Our current site just lists the "watch" and "stars" values from a package's GitHub repository. This data doesn't present any useful information since it's very dependent on the long-term history of a package, and newer packages are unlikely to be able to "break into" the top 5 packages even if they are suddenly very possible.
- Provide documentation of architecture and processes so that the site can be maintained by the Zeek team going forward

Looking at other existing package sites is a great way to draw inspiration about what can be done. One site the team likes in general is <https://crates.io>, which is the package site for Ruby.

Knowledge, Skills, and Expertise Required:

- Web programming skills (frontend and backend)
- Working knowledge of git

Equipment Required:

- No significant equipment required, other than the ability to host the site

Software and Other Deliverables:

- A working prototype of the site
- All code posted to GitHub under the BSD license