1. Overview

Motivation for the Data Collection module is that designers, researchers, and Argus users need to record certain information about their use of Argus. For example, how often users create new rules or rulesets, how often they delete or change them, how often a given ruleset was invoked, and how often a given rule fired on a message. Other types of relevant data include information on the different *types* of messages being processed.

The design of the data collection module allows different categories of data to be collected independently of each other. For example, one can collect data about rule creation and deletion without collecting information about rule firing activity. There are three types of data collected by the data module: rule activity, rule life, and message activity. Rule activity data logs testing and firing information about rules. Rule life data logs creation, deletion, or changing of rules and rulesets. Message activity data logs information about the processing of messages.

Data for a given user is stored (by default) in files in the *data* subdirectory of that user's *lens* directory. The default names for the three files are as follows:

- *lens.rl.log*: stores data about creation, changing, and deletion of rules and rulesets.
- *lens.ra.log*: stores information for every rule in each rule set; the information shows how many messages a rule was tested on and, of those, how many the rule fired on.
- *msg.log*: stores data about messages that are processed by Argus, such as the *type* of the message and the time at which it was processed.

2. Rule Life Data

Each time a rule or ruleset is created, deleted, or modified, the *lens.rl.log* file records the event. A ruleset is considered to be modified when a rule is added to or removed from the ruleset, or when a rule in the ruleset is modified. The log file refers to rules and rulesets by the date at which they were last changed. Rulesets are also identified by their name, whereas for rules the names are optional.

When a rule is created the rule itself is recorded in the log file, along with its date identifier.

3. Message Activity

Each time Argus handles a message the *msg.log* records the time of the event and the type of message processed.

4. Rule Activity Data

Unlike with rule life and message activity logging, information about rule activity is not immediately recorded in a log file. The data collection module keeps track of rule activity over a certain period of time (which the user modify). Then, at the end of the time period, statistics are recorded in *lens.ra.log* for each rule in every ruleset for that period. The statistics show how many messages the rule was tested on and how many the rule fired on.

5. Formats of the Data Collection Files

There are four files associated with data collection. The three files listed above are in a human readable format. The fourth file, *lens.data*, is read from and written to by the data collection module and should not be modified. It is, however, an ascii file with the following format:

```
format_version <num>        /* format version number */
**RULE-LIFE:
<enabled>                    /* indicated whether specific collection ok */
<in-time>                   /* used internally for computation */
<start-time>                /* format: seconds since Jan 1, 1970 */
<stop-time>                 /* same format */
<period>                     /* collection period length */
**RULE-ACTV:
<enabled>                    /* indicated whether specific collection ok */
<in-time>                   /* used internally for computation */
<start-time>                /* format: seconds since Jan 1, 1970 */
<stop-time>                 /* same format */
<period>                     /* collection period length */
**MSG:
<enabled>                    /* indicated whether specific collection ok */
<in-time>                   /* used internally for computation */
<start-time>                /* format: seconds since Jan 1, 1970 */
<stop-time>                 /* same format */
<period>                     /* collection period length */

/* the following occur only if rule-activity statistics are still being
   collected at the time of termination of rule running
*/
<rule-date-id> #<num>   /* same format; number of rule sets following */
<rule-set-name> #<num>  /* number of message type following */
<message-type-name> <num> <num> <num> <num> <num> <num>
        .
        .
        .
/* number of tests, then fires for shuffled, incoming,
   and outgoing messages
*/
```

So a sample *lens.data* file might look like this:

```
format_version 5
**RULE-LIFE:
1
1
614620707
630172707
15552000
**RULE-ACTV:
1
1
614829469
614833069
3600
**MSG:
1
1
```

```
2

614620716
630172716
15552000

614554829 #1
rset #2
phone 4 0 30 5 0 0
meeting 2 2 0 0 0 0
```

The block of four lines starting at the line break are temporary. They contain information about the statistics of rule testing and firing being collected in the current rule activity collection period. When the collection period has ended the block of data is deleted here and put in a more human-readable form in the file *lens.ra.log*. The above block records that within the current collection period, the rule created on Thursday, June 22 1989, at 17:40:29, was invoked by one ruleset, "rset". The rule was tested on two different message types, "phone", and "meeting". For the "phone" messages, the rule was tested on 4 and fired on none while shuffling messages; it was tested on 30 and fired on 5 incoming messages; it was tested on no outgoing messages.

Table of Contents