R: A Proposed Analysis and Visualization Environment for Network Security Data

Josh McNutt <jmcnutt@cert.org>

Outline

SiLK Tools

Analyst's Desktop

Introduction to R

R-SiLK Library (Proof-of-concept prototype)

Context Objects and Analysis Objects

Analyst Benefits

Prototype Demo

Future of Analyst's Desktop

CERT

SiLK Tools

System for Internet-Level Knowledge

- http://silktools.sourceforge.net/

Developed and maintained by CERT/NetSA (Network Situational Awareness) Team

Consists of a suite of tools which collect and perform analysis operations on NetFlow data

Optimized for very large volume networks

Command Line Interface (CLI)

Fundamental Tools

- rwfilter
- rwcount
- rwuniq
- IP sets

CERT

Enhancing SiLK: Analyst's Desktop

We are currently developing a new interface model for the SiLK tools

The goal is to develop an environment supporting sophisticated analysis of network security phenomena

- **Analyst's Desktop**

Requirements

- Interactive visualization capability
- Audit trail
- Annotation
- Preserve the command line options available in SiLK
- Make simple analyses simple to perform

Platform of choice: **R**

CERT

R: What is it?

R is a programming language and environment for statistical computing and graphics used by statisticians worldwide

The R Project for Statistical Computing

- http://www.r-project.org

R is available as Free Software under the terms of the Free Software Foundation's GNU General Public License in source code form

There exists a thriving community of statisticians and statistical programmers who contribute their code

CERT

R! What is it good for?

R represents "best-in-practice" environment for exploratory data analysis

Specifically designed with data analysis in mind

- A more natural analysis interface than Perl, Python or shell scripts

Full Access to R's built-in statistical analysis capability

R can run interactively or in batch mode

Visualization

- Integrated graphing capabilities (publication quality)

CERT

R! What is it good for?

Object-based environment

- Everything in R is an object
  - functions, matrices, vectors, arrays, lists
- Objects can be saved in user workspace (persistence) or saved to disk and sent to another user's workspace
- Preserve results for comparison with future analyses
- Annotations can be attached to objects

Command line control can be preserved

- Wrapper functions incorporate SiLK command line arguments

Rapid prototyping of new analysis techniques and visualizations
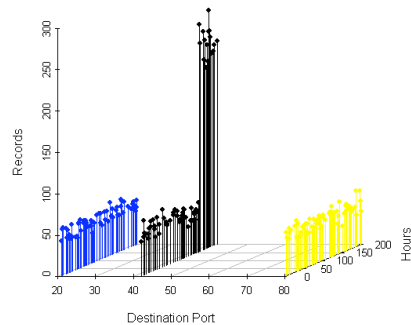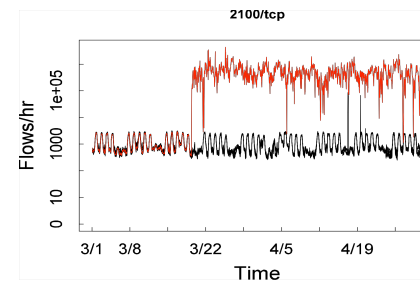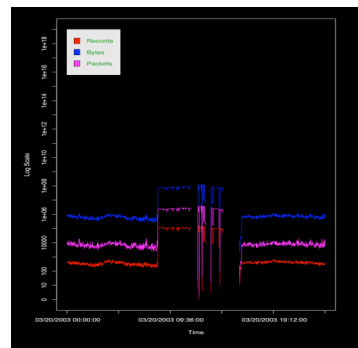
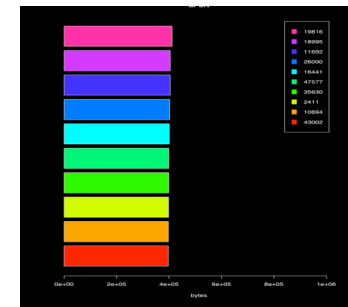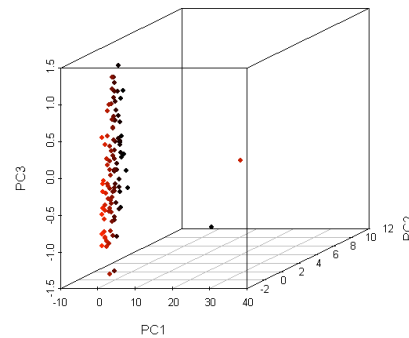CERT

R's Graphing Capability

Huge set of standard statistical graphs

- stemplots, boxplots, scatterplots, etc.

3D graphing capability

CERT

R-SiLK Library

Low-level interface involves custom wrapper functions making
command line calls to SiLK

Higher-level functions call those wrappers

Many SiLK Tools have associated functions in R-SiLK library

- rwfilter(), rwcount(), rwuniq(), rwcut()

The SiLK Tool "rwcount" generates a binned time series of
records, bytes and packets

In R-SiLK library, there is a function called "rwcount()"

- rwcount(rwcount switches, context object)
- Example below assigns 60-second binned time series data
  for context object called "context.tcp" to the analysis object
  called "analysis.tcp"
  - analysis.tcp <- rwcount("--bin-size=60", context.tcp)

Context objects and analysis objects?

CERT

Context Objects and Analysis Objects

To aid in analysis tasks, we've created something called a context object

Context Object

- An object in R that determines precisely what data is being considered
  - Contains a text string element indicating filter criteria (**rwfilter switches**)
  - Contains the name of the binary file of flow data satisfying the filter criteria
- Simple example (Time period is only filter criteria)
  - All flows between midnight and 1 a.m. on July 17th, 2005
- Advanced example (Many additional criteria)
  - All inbound flows from source XXX.YYY.XXX.ZZ between midnight and 1 a.m. on July 17th, 2005 targeting any hosts in XXX.ZZZ.0.0/16 on destination port 42/tcp

As the analysts learn more about a particular context through analysis, they will be able to refine the current context by adding additional filter criteria

CERT

Context Objects and Analysis Objects

Context objects can be summarized/described via the process of analysis
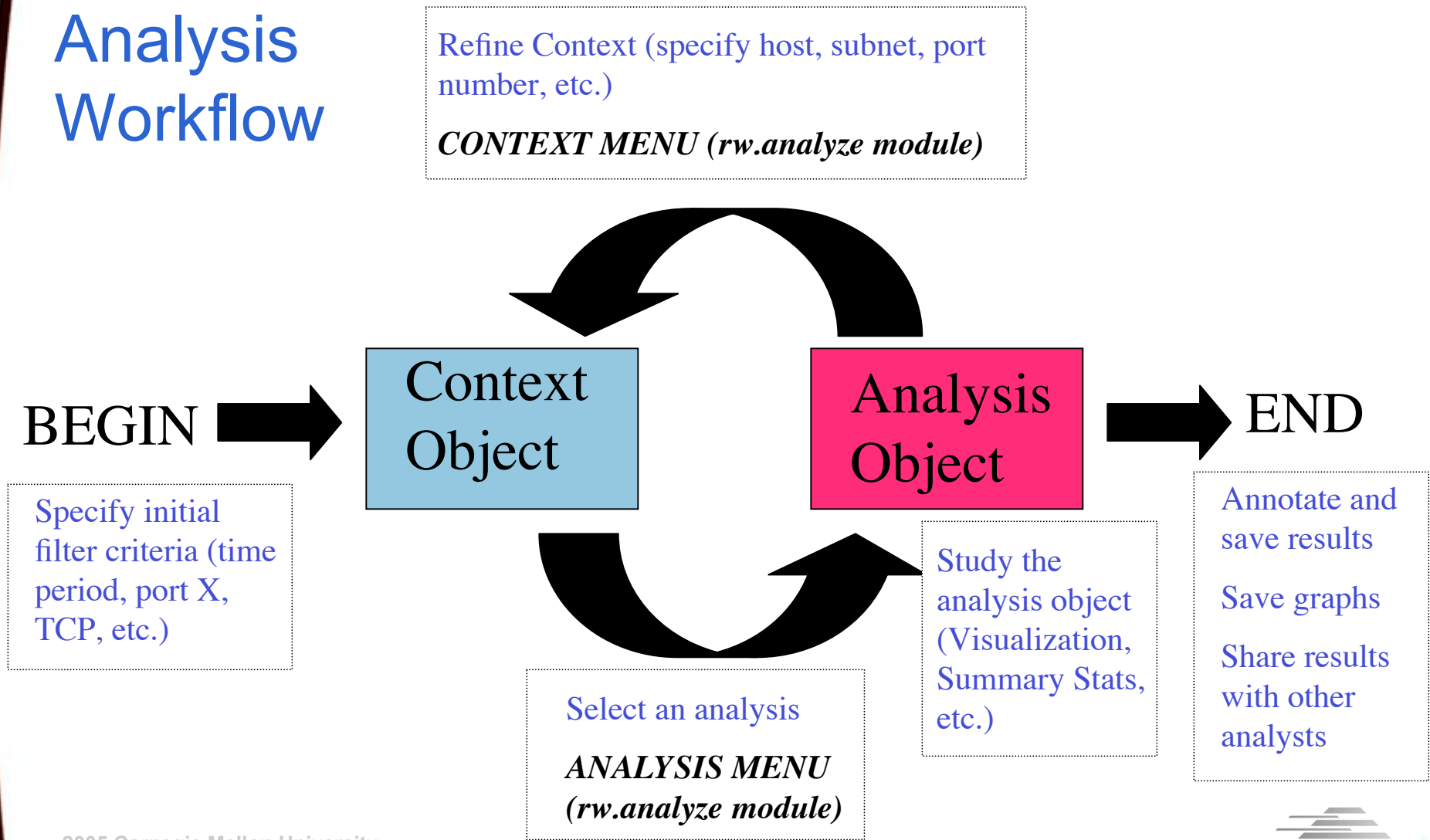
To store the results of analysis we have analysis objects

Analysis Object

- An object in R that saves a description of a context object
  - Examples:
    - A top N list of destination ports for a context object
    - A binned time series of the flow data for a context object
- Components
  - Data (time series, sorted list of port volumes, etc.)
  - Context Object (what was the source data)
  - Timestamp (when was it created)
  - Descriptive Results (correlation, mean, etc.)
- Annotation
  - Can be attached to analysis object by analyst
  - Examples:
    - "UDP-based DDoS began around 8:30 a.m. on 5/6/04"
    - "Scanning appears to be targeting 2 local subnets"

CERT

Context Objects and Analysis Objects

Analysis Workflow

Refine Context (specify host, subnet, port number, etc.)

CONTEXT MENU (rw.analyze module)

BEGIN → **Context Object** ⟳ **Analysis Object** → END

Specify initial filter criteria (time period, port X, TCP, etc.)

Select an analysis

ANALYSIS MENU (rw.analyze module)

Study the analysis object (Visualization, Summary Stats, etc.)

Annotate and save results

Save graphs

Share results with other analysts

CERT

Analyst Benefits

Experienced Analyst

- Enhanced command line experience
  - Immediate and integrated visualization
- Object Persistence
- Annotation
- Audit Trail
- Rapid Prototyping

Beginner Analyst

- Faster time to productive investigations
- rw switches can be made transparent to the user
  - concatenated together in the background
- rw.analyze() module

Prototype Demo

R interactive mode

Basic proof-of-concept interface: rw.analyze()

Demonstrate the *Context Object – Analysis Object* workflow

Begin Demo

CERT

Future of Analyst's Desktop

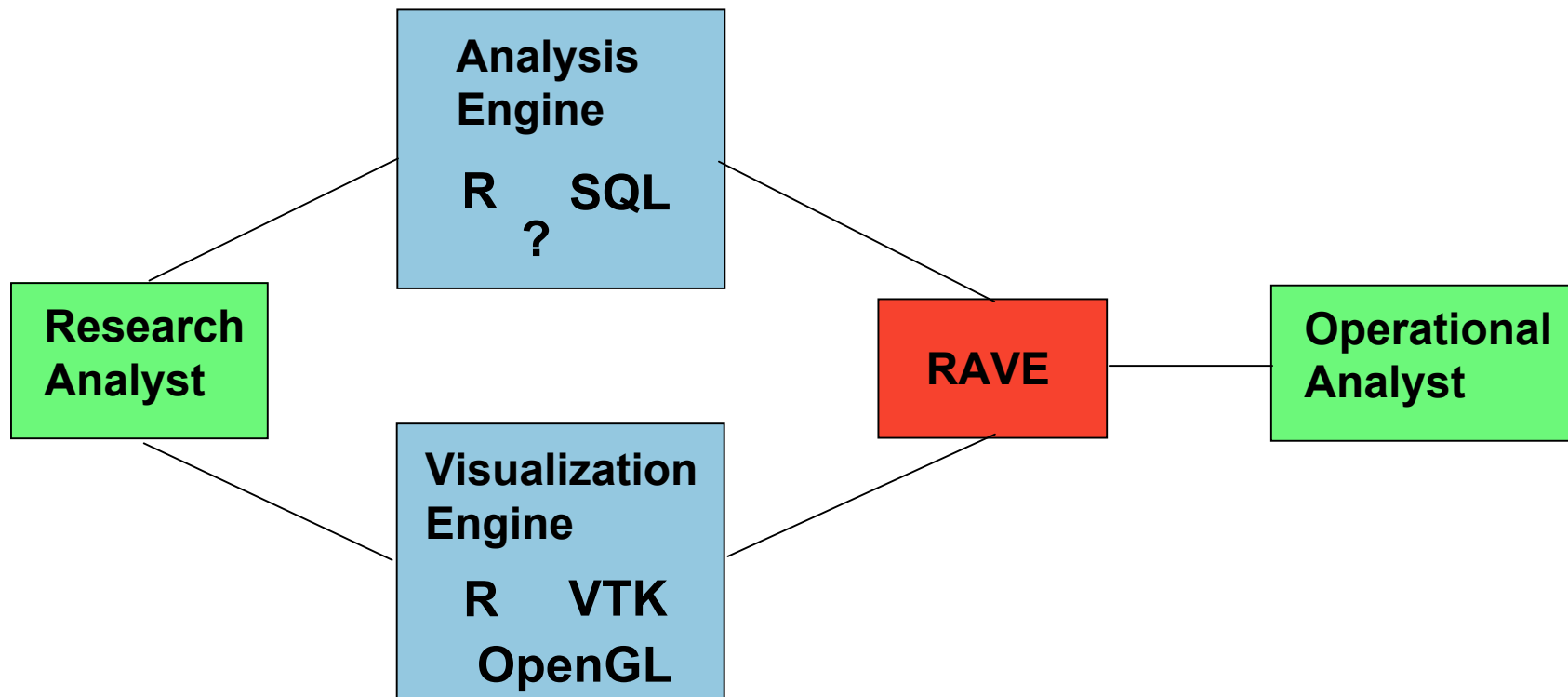Working on improved version of R-SiLK library and prototype interface

Support different modes of analysis

- Research Analysis
  - Flexible, powerful, customized
- Operational Analysis
  - Immediate, concise, "canned"

CERT

Future of Analyst's Desktop

RAVE

- Retrospective Analysis and Visualization Engine

Future of Analyst's Desktop

RAVE

- **Operationalize analysis techniques**
  - Move new research techniques efficiently into operations
  - Furnish operational services (e.g. caching)
- **Decouple analysis/visualization from UI**
  - Different A/V tools, same UI
    - SiLK, R, SQL, Python/C, etc.
  - Different UIs, same engine
    - "Dashboards"
    - Menu of "canned" queries
    - Sophisticated data exploration environment (e.g., R)

CERT

Questions

CERT