Visualization is Better!  
A Comparative Evaluation

John Goodall
johng@securedecisions.avi.com
Secure Decisions division of Applied Visions, Inc.
Context

• This work was part of larger research study
• Field study, interviews with security analysts, and survey to understand intrusion detection work practice
• Development of vis tool for analysis
  – Iterative heuristic reviews and usability testing
• Summative comparative evaluation
User Testing

• Controlled experiments comparing design elements: a comparison of specific widgets

• Usability evaluation of a tool: an evaluation of problems users encounter when using a tool as part of the design process

• Controlled experiments comparing two or more tools: a comparison of multiple visualizations or the state of the art with a novel visualization

• Case studies of tools in realistic settings: an evaluation of a visualization tool in a natural setting with users using the tool to accomplish real tasks
User Testing

• Controlled experiments comparing design elements: a comparison of specific widgets

• Usability evaluation of a tool: an evaluation of problems users encounter when using a tool as part of the design process

• Controlled experiments comparing two or more tools: a comparison of multiple visualizations or the state of the art with a novel visualization

• Case studies of tools in realistic settings: an evaluation of a visualization tool in a natural setting with users using the tool to accomplish real tasks
Study Design

- Goal: Compare tnv and the standard tool for network packet analysis
- Design: Repeated measure within subject
- Participants: 8 IS undergrad/grad students
- Tools: tnv & Ethereal
- Data: small (200 packets) & large (750 packets)
- Tasks: well-defined & exploratory
Why Novice Users?

• Learning: research showed that novices ‘play’ with tools to learn; tnv was designed to facilitate learning

• Background: domain experts would have lots of experience with Ethereal, which could skew the results

• Accessibility: domain experts are hard to come by
Tools

- Wireshark
- tnv

**Wireshark**

*De facto standard for packet analysis: 88% of survey respondents used Ethereal at least occasionally (62% frequently)*

**tnv**

*Designed to facilitate high-level and detailed understanding of network traffic*
Tasks

• Well-defined
  – Representative of ‘typical’ tasks; 1 correct answer
  – Task categories: comparison & identification
  – 16 tasks for each tool

• Exploratory
  – Asked participants to draw open ended conclusions from the data; no correct answer
  – Predefined time limit
  – 1 exploratory task for each tool
Tasks

• **Well-defined**
  – Representative of ‘typical’ tasks; 1 correct answer
  – Task categories: comparison & identification
  – 16 tasks for each tool

• **Exploratory**
  – Asked participants to draw open ended conclusions from the data; no correct answer
  – Predefined time limit
  – 1 exploratory task for each tool
Procedure

• Introduction to the study and each of the tools
• Training using either tnv or Ethereal
• Timed tasks using that tool
• Exploratory task using that tool
• Training using the second tool
• Timed tasks using the second tool
• Exploratory task using the second tool
• A satisfaction questionnaire on both tools
Variables

- **Independent Variables**
  - Tool: tnv, Ethereal
  - Task Type: Comparison, Identification

- **Dependent Variables**
  - Accuracy
  - Completion Time
  - User Perceptions
Expected Results

Expect users to perform better with tnv...

...Especially for comparison tasks, since tnv shows much more data at once

...But identification tasks will be closer, since Ethereal has easy to use search capability
Analysis

• A repeated measures analysis of variance (RMANOVA) with repeated measures for tool (tnv, Ethereal) and task type (Comparison, Identification)

• To ensure that counterbalancing the tool order usage had no effect on performance, order was treated as a between subject variable

• The between subject variable of tool order was not significant in any of the tests
Interaction effect of tool: \( F(1,6) = 14.72, p = 0.009 \)

Participants had significantly fewer errors using tnv than using Ethereal.
Accuracy

Interaction effect between tool and task type:
\[ F(1,6) = 2.139, \ p = 0.194 \]

But, looking at comparison tasks for each tool, there is an effect
\[ t = 5.612, \ p = 0.001 \]

Mean and 95% confidence interval of accurate responses by tool and task type. (max. = 5)
Time

• Time to completion for *successful* tasks
  – Not partially successful tasks or timed out tasks
  – Incorrect responses could have been guesses

• Standardized time
  – Tasks were of varying levels of difficulty
  – Average time for each task varied greatly
  – Negative number means faster than average

\[
\text{StandardizedTime} = \frac{(ParticipantTime - TaskMeanTime)}{TaskStandardDeviation}
\]
Interaction effect of tool: $F(1,6) = 5.581, p = 0.056$

Trend suggests faster performance, but not significant
Time

Interaction effect between tool and task type
\( F(1,6) = 2.558, p = 0.161 \)

But, looking at comparison tasks for each tool, there is an effect
\( t = -4.615, p = 0.002 \)

Mean and 95% confidence interval of standardized time to successful tasks by tool and task type
Discussion: Task Type

• Larger difference in comparison tasks
  – Ethereal: Statistics were underused; comparisons were done by sorting and mental addition
  – tnv: Comparisons could be seen at a glance

• Less of a difference in identification tasks
  – Ethereal: Search on small data sets removed all but the relevant information
  – tnv: Search highlighted relevant information, but kept all data on the screen, so participants didn’t always see where it was
Discussion: Tasks

Accuracy by Task

Proportion of correct responses

Task
Port Related Tasks

- Tasks 2, 3: compare port activity
- tnv port visualization is hidden by default
- Participants couldn’t answer by looking at main display
- Participants learned in task 2, so task 3 was much faster (81 s -> 22 s)
Exploratory Tasks

• Measured number of ‘insights’ that were not mentioned in timed tasks and not incorrect
• Results: participants often started out talking about the tools, not the data
• Several simply gave up (especially for Ethereal)
Results: Exploration

- tnv: higher-level
  - Gap in activity
- Ethereal: packet-level details
  - Unencrypted passwords

Mean and 95% confidence interval of the number of insights discovered
User Perceptions

Satisfaction Ratings by Tool

- TNV
- Ethereal

Mean Score vs. Satisfaction

- Overall satisfaction
- Ease of processing information
- Ease of searching
- Ease of learning
- Ease of seeing patterns
- Level of confidence
- Level of performance
Ease of Seeing Patterns

Satisfaction Ratings by Tool

Mean Score

Satisfaction

Overall satisfaction  Ease of processing information  Ease of searching  Ease of learning  Ease of seeing patterns  Level of confidence  Level of performance

TNV

Ethereal
Lessons

• Domain experts are difficult to recruit
  – Include them in the design process
• Training can take a lot of test time
  – Self-directed training matches how analysts learn
• Data sets are problematic and unlabeled
• ‘Realistic’ tasks that can be answered quickly
  with both tools are hard to define
  – ???
Questions?

John Goodall
johng@securedecisions.avi.com
Secure Decisions division of Applied Visions, Inc.