Supporting the Cyber Analytic Process using Visual History on Large Displays

Ankit Singh, Alex Endert, Christopher Andrews, Lauren Bradel, Robert Kincaid, Chris North

Virginia Tech

Agilent Laboratories

Overview

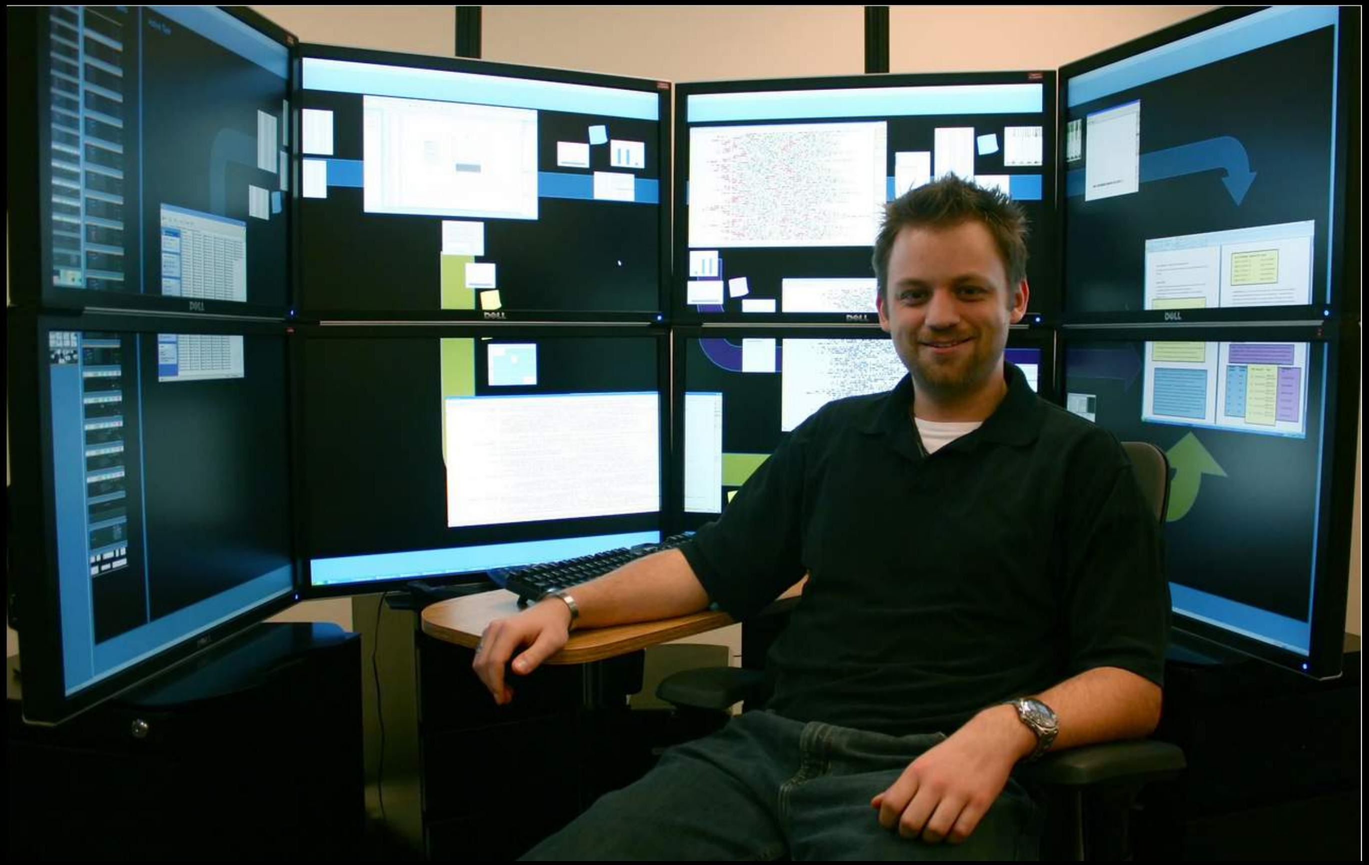
- Cyber Analytic Process
  - Benefits provide by large displays
- Visual History Design and Prototype
- Lessons Learned, Future Work

Large, High-Resolution Displays

- Personal Workspace
- Single Workstation
- Familiar OS, tools, ...
- Provides additional size, resolution to support analysts

Cyber Analytic Process

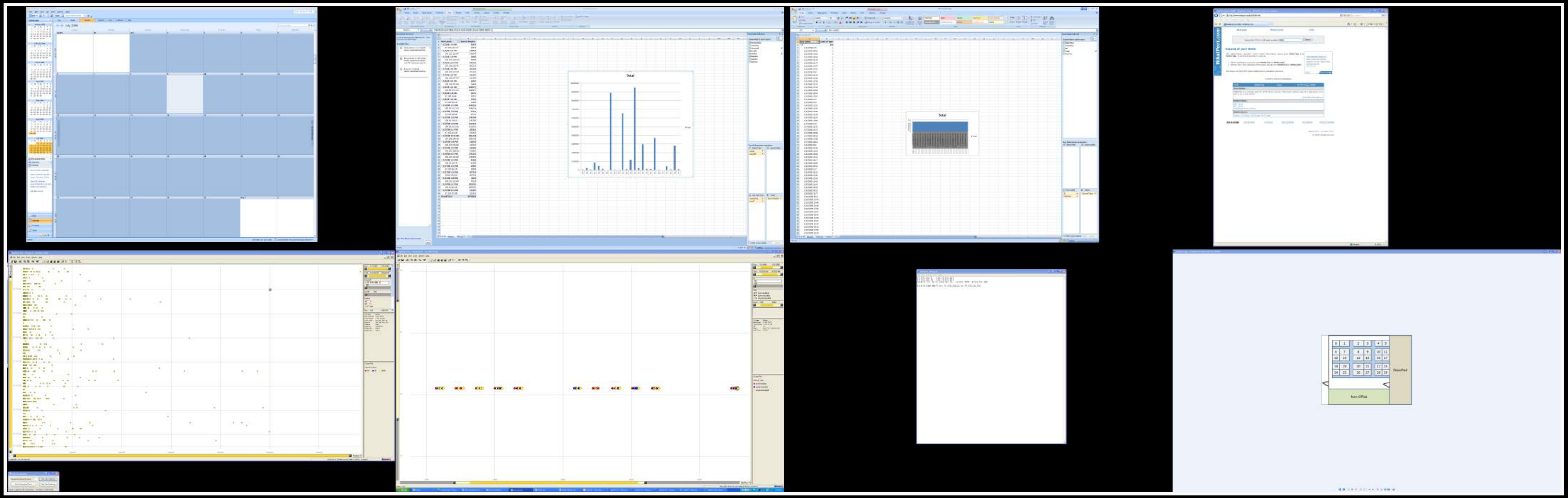
- Interviewed 8 professional cyber analysts
- Observed 4 analysts analyze the 2009 VAST Challenge Dataset
  - Simulated Network Flows and Employee Building Access logs



from Fink, G., North, C., Endert, A. and Rose, S. Visualizing Cyber Security: Usable Workspaces. VizSec, 2009.

Cyber Analytic Process

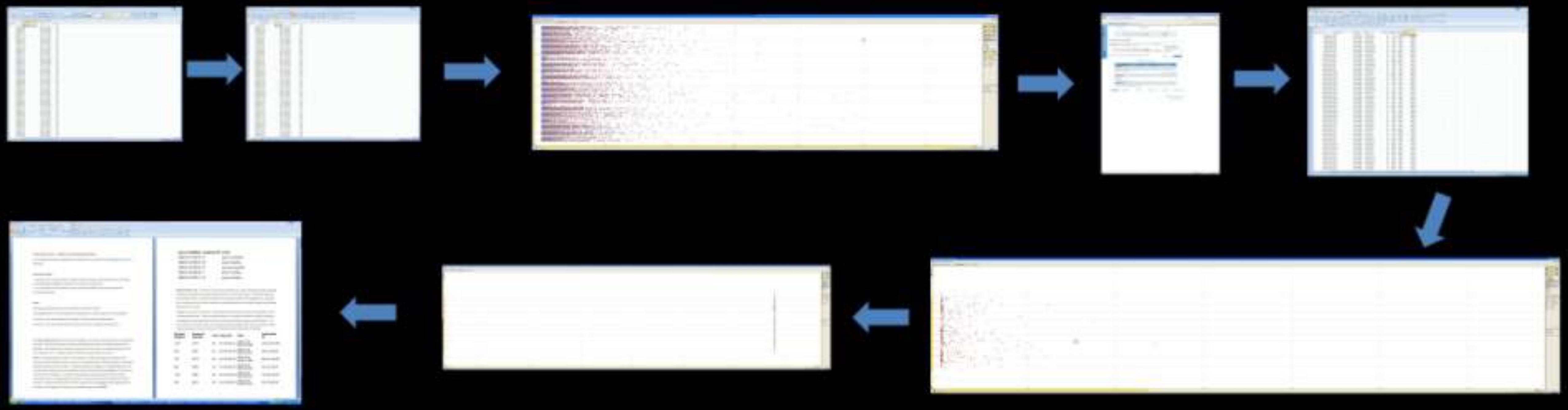
- Multiple data sources
- Multiple tools/windows
- Extensive Excel usage



from Fink, G., North, C., Endert, A. and Rose, S. Visualizing Cyber Security: Usable Workspaces. VizSec, 2009.

Cyber Analytic Process

- Versioning of files based on hypotheses
  - E.g., v1.1, v1.2, v2.1, ...
  - Reasons: save the data, save the view
- Difficult to re-create process to support findings at time of creating report



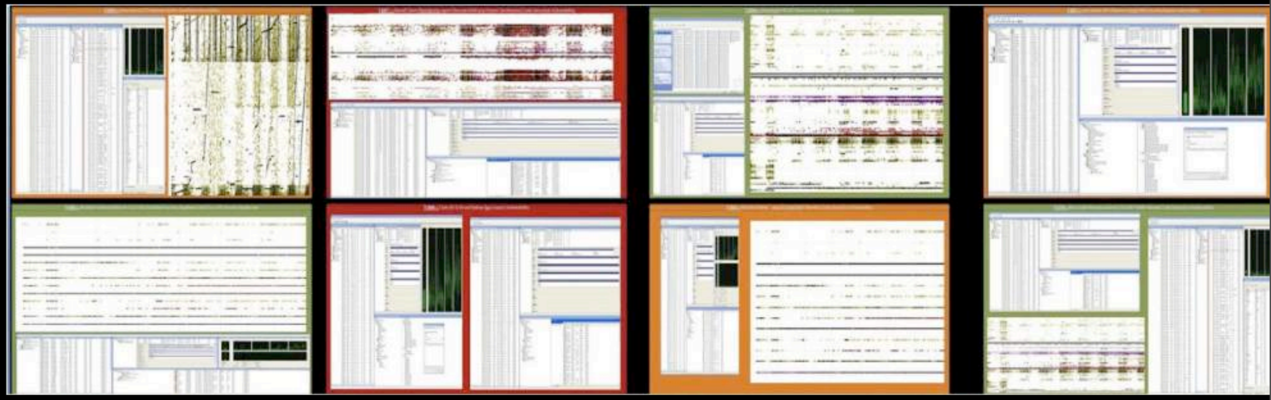
from Fink, G., North, C., Endert, A. and Rose, S. Visualizing Cyber Security: Usable Workspaces. VizSec, 2009.

Challenge

- How to design workspaces to support the complex cyber analytic process?

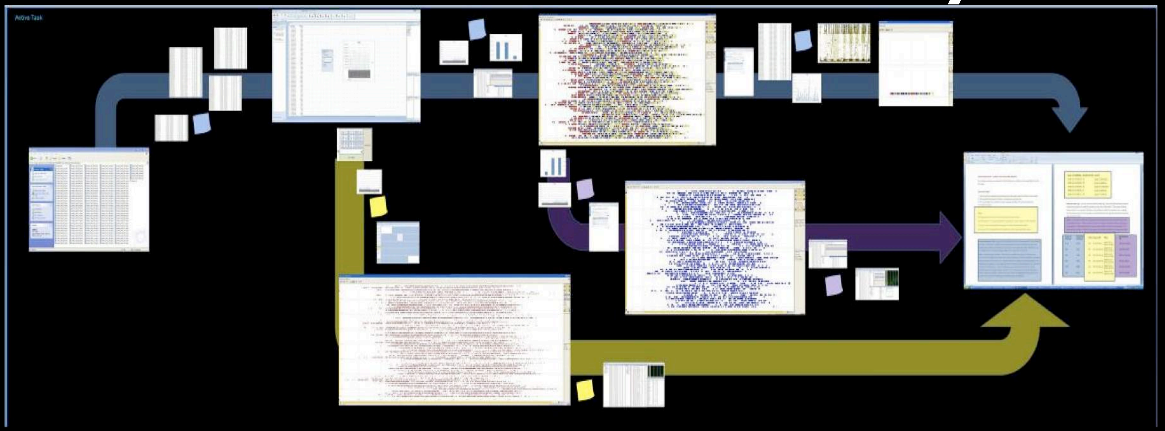
More Resolution and Size

De-Aggregation of Data



Case Management

Process History



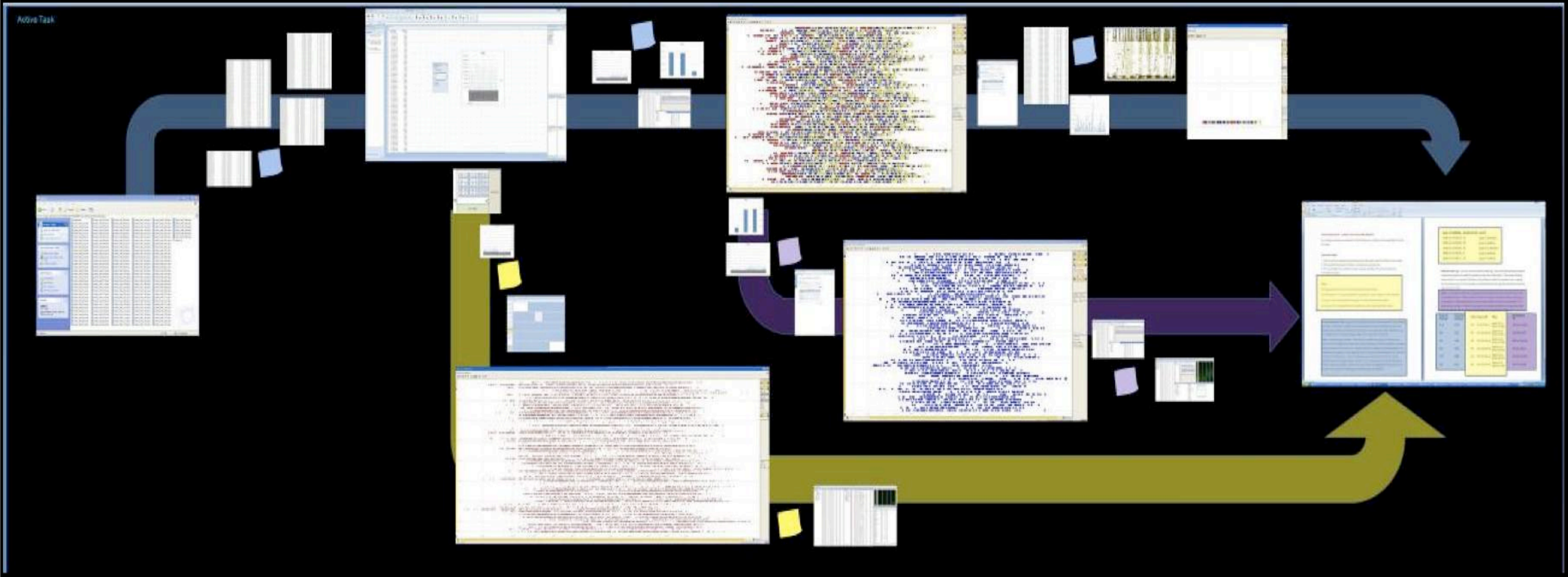
from Fink, G., North, C., Endert, A. and Rose, S. Visualizing Cyber Security: Usable Workspaces. VizSec, 2009.

Visual History: Design

Branching

Multiple Windows, File Versions

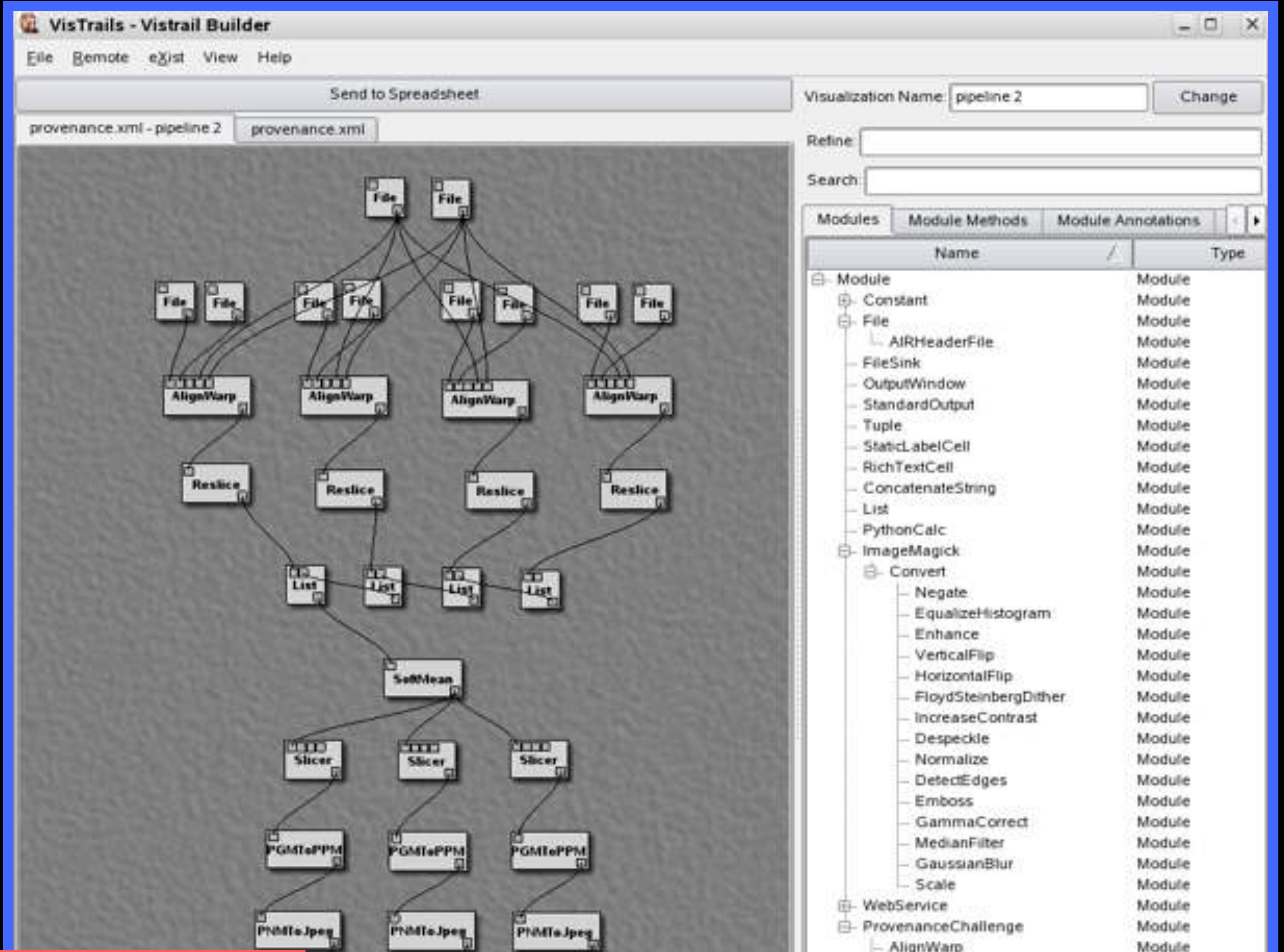
Process Traceability



from Fink, G., North, C., Endert, A. and Rose, S. Visualizing Cyber Security: Usable Workspaces. VizSec, 2009.

Visual History: Design

Visualization of workflow



VisTrails - Vistrail Builder

File Remote eXist View Help

Send to Spreadsheet

provenance.xml - pipeline 2 | provenance.xml

Visualization Name: pipeline 2 | Change

Refine:

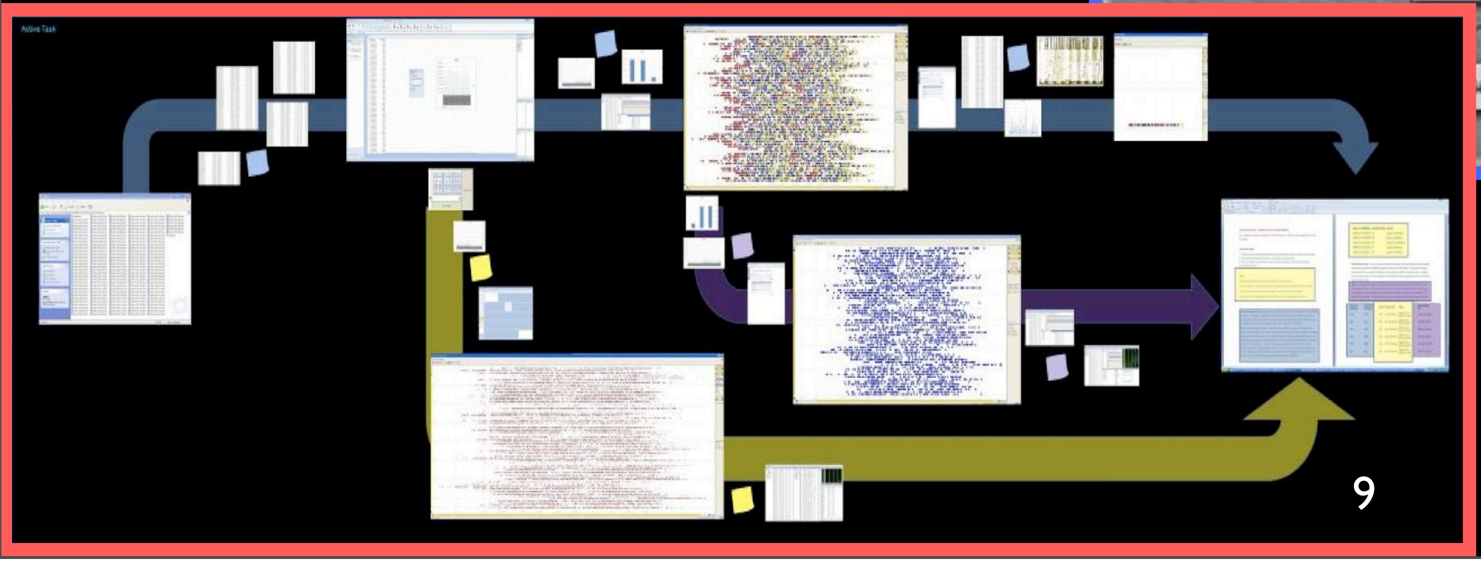
Search:

Modules | Module Methods | Module Annotations

Name	Type
Module	Module
Constant	Module
File	Module
AIRHeaderFile	Module
FileSink	Module
OutputWindow	Module
StandardOutput	Module
Tuple	Module
StaticLabelCell	Module
RichTextCell	Module
ConcatenateString	Module
List	Module
PythonCalc	Module
ImageMagick	Module
Convert	Module
Negate	Module
EqualizeHistogram	Module
Enhance	Module
VerticalFlip	Module
HorizontalFlip	Module
FloydSteinbergDither	Module
IncreaseContrast	Module
Despeckle	Module
Normalize	Module
DetectEdges	Module
Emboss	Module
GammaCorrect	Module
MedianFilter	Module
GaussianBlur	Module
Scale	Module
WebService	Module
ProvenanceChallenge	Module
AlignWarp	Module
Reslice	Module
SoftMean	Module
Slicer	Module
PGMToPPM	Module
PNMToJpeg	Module

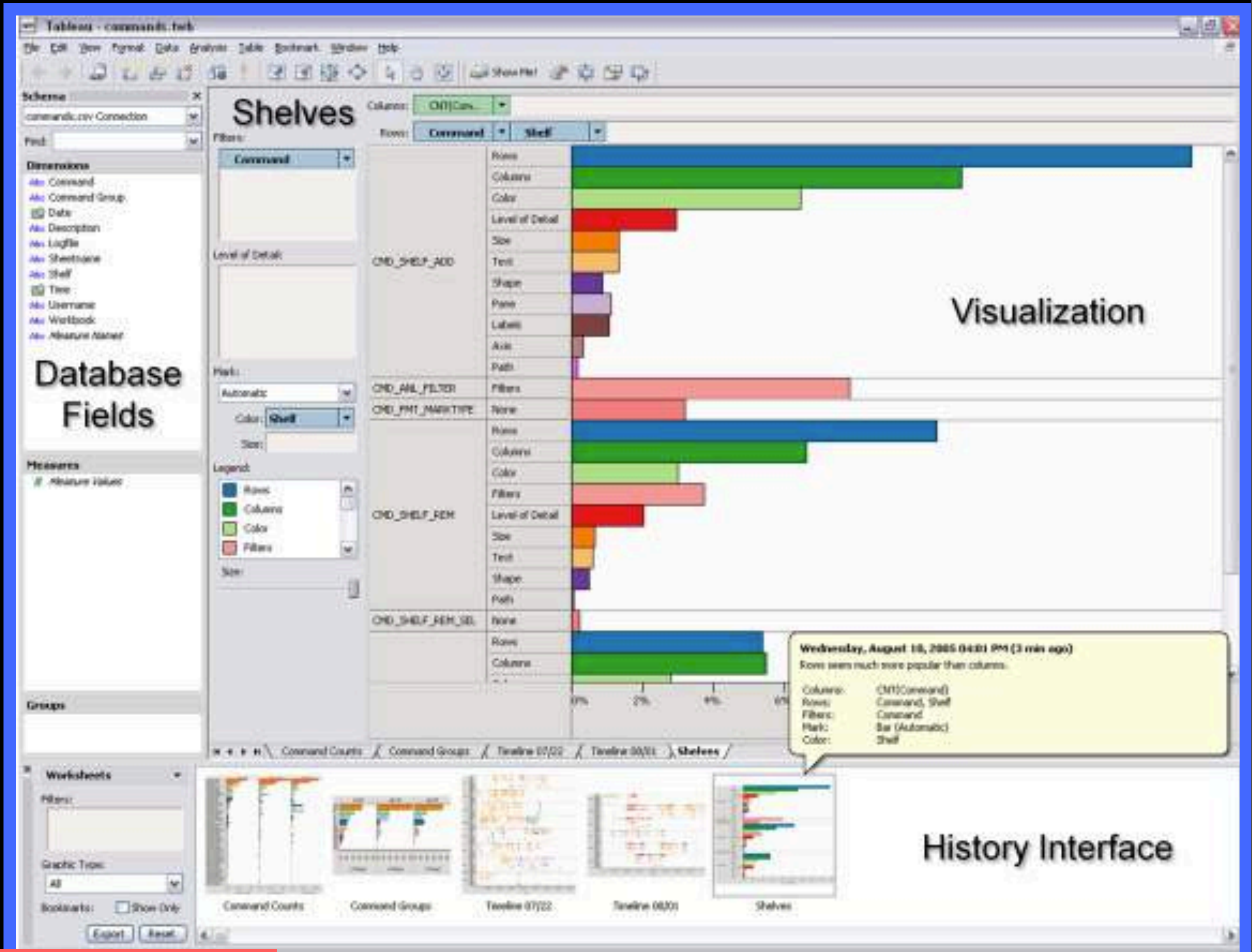
VisTrails, http://www.sci.utah.edu/~vgc/vistrails

Process integrated in workspace



Visual History: Design

History stored away in thumbnails

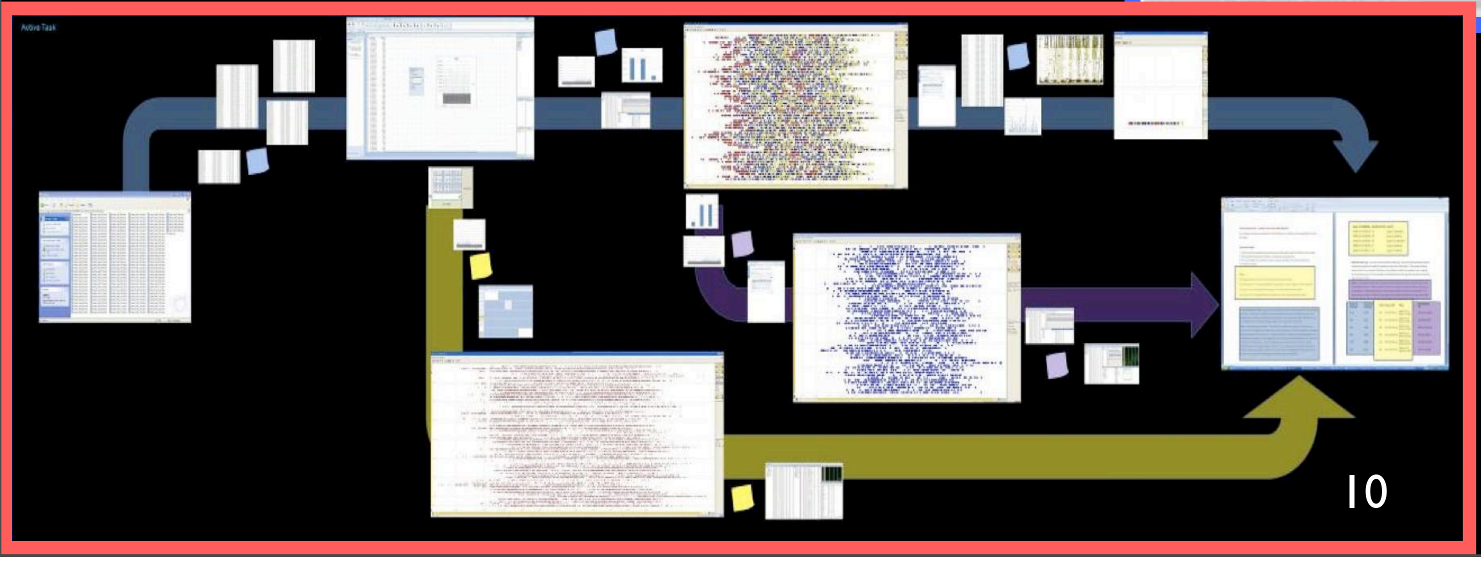


Wednesday, August 10, 2005 04:01 PM (3 min ago)
Rows seem much more popular than columns.

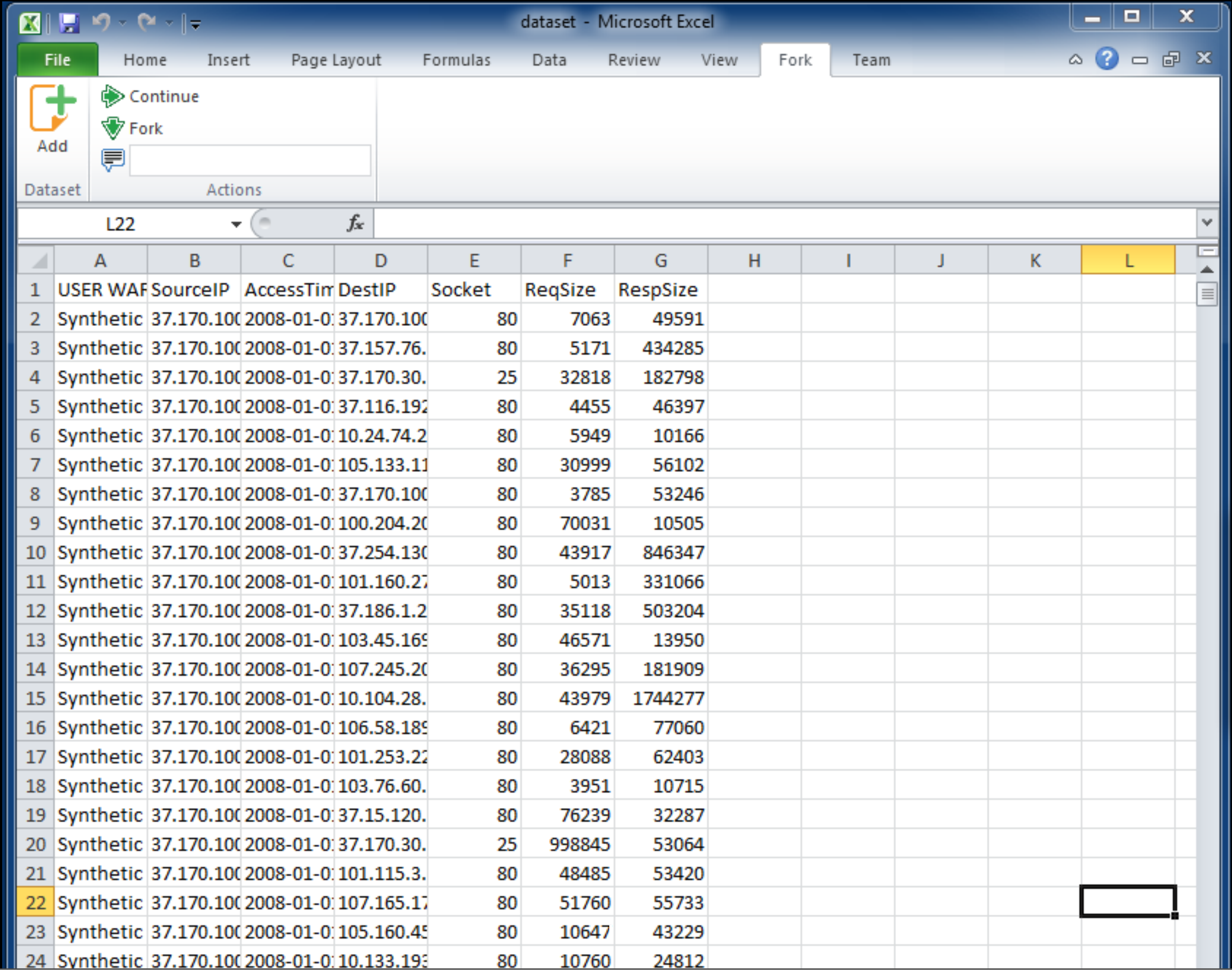
Columns:	CNT(Command)
Rows:	Command, Shelf
Filters:	Command
Mark:	Bar (Automatic)
Color:	Shelf

Tableau, image from http://hci.stanford.edu/jheer/files/jheer-thesis.pdf

Process integrated in workspace



Visual History: Implementation



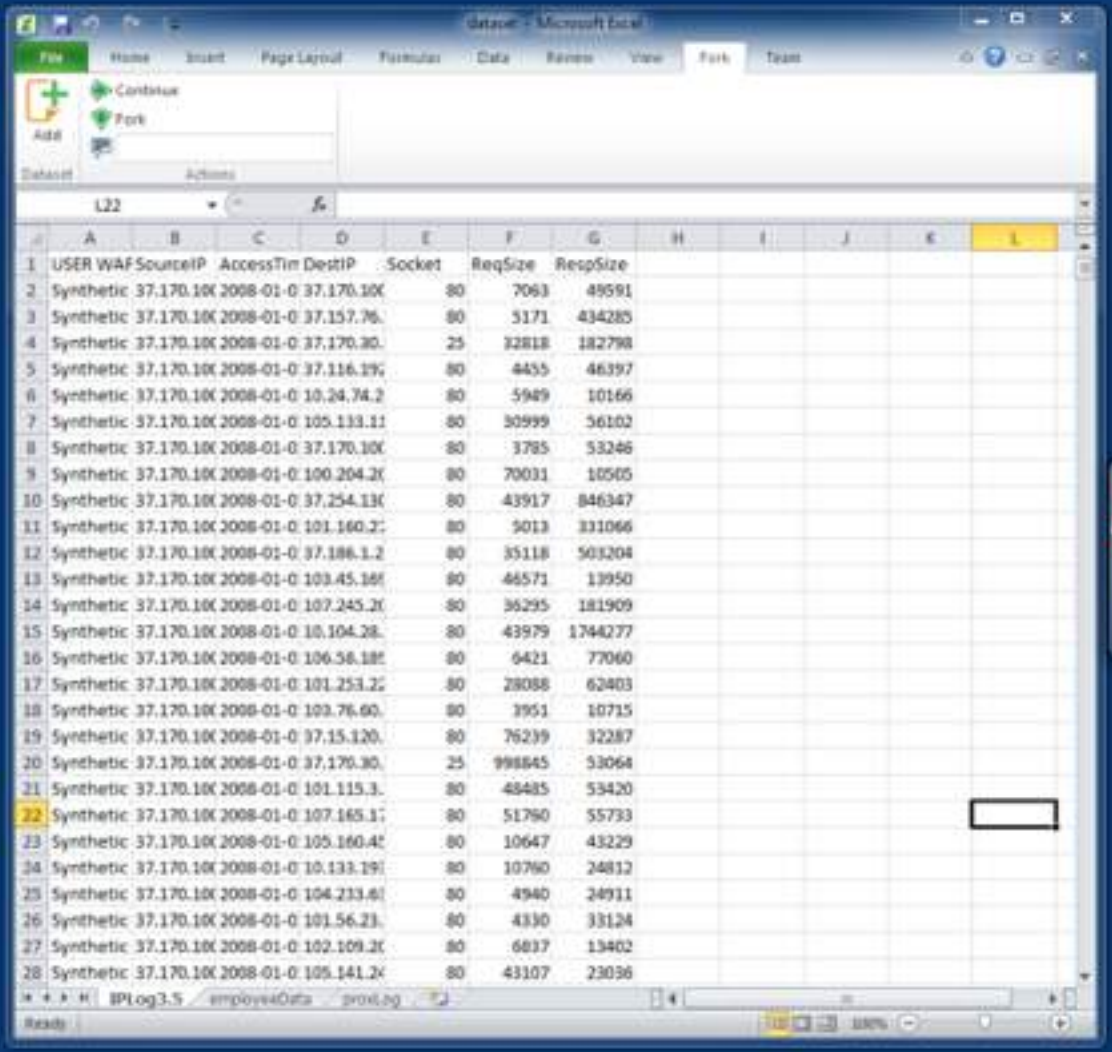
	A	B	C	D	E	F	G	H	I	J	K	L
1	USER WAF	SourceIP	AccessTim	DestIP	Socket	ReqSize	RespSize					
2	Synthetic	37.170.100	2008-01-0	37.170.100	80	7063	49591					
3	Synthetic	37.170.100	2008-01-0	37.157.76.	80	5171	434285					
4	Synthetic	37.170.100	2008-01-0	37.170.30.	25	32818	182798					
5	Synthetic	37.170.100	2008-01-0	37.116.192	80	4455	46397					
6	Synthetic	37.170.100	2008-01-0	10.24.74.2	80	5949	10166					
7	Synthetic	37.170.100	2008-01-0	105.133.11	80	30999	56102					
8	Synthetic	37.170.100	2008-01-0	37.170.100	80	3785	53246					
9	Synthetic	37.170.100	2008-01-0	100.204.20	80	70031	10505					
10	Synthetic	37.170.100	2008-01-0	37.254.130	80	43917	846347					
11	Synthetic	37.170.100	2008-01-0	101.160.27	80	5013	331066					
12	Synthetic	37.170.100	2008-01-0	37.186.1.2	80	35118	503204					
13	Synthetic	37.170.100	2008-01-0	103.45.169	80	46571	13950					
14	Synthetic	37.170.100	2008-01-0	107.245.20	80	36295	181909					
15	Synthetic	37.170.100	2008-01-0	10.104.28.	80	43979	1744277					
16	Synthetic	37.170.100	2008-01-0	106.58.189	80	6421	77060					
17	Synthetic	37.170.100	2008-01-0	101.253.22	80	28088	62403					
18	Synthetic	37.170.100	2008-01-0	103.76.60.	80	3951	10715					
19	Synthetic	37.170.100	2008-01-0	37.15.120.	80	76239	32287					
20	Synthetic	37.170.100	2008-01-0	37.170.30.	25	998845	53064					
21	Synthetic	37.170.100	2008-01-0	101.115.3.	80	48485	53420					
22	Synthetic	37.170.100	2008-01-0	107.165.17	80	51760	55733					
23	Synthetic	37.170.100	2008-01-0	105.160.45	80	10647	43229					
24	Synthetic	37.170.100	2008-01-0	10.133.193	80	10760	24812					

Visual History: Implementation

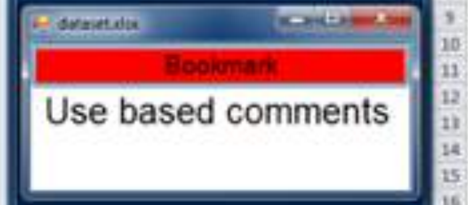
Branching

Multiple Windows,
File Versions

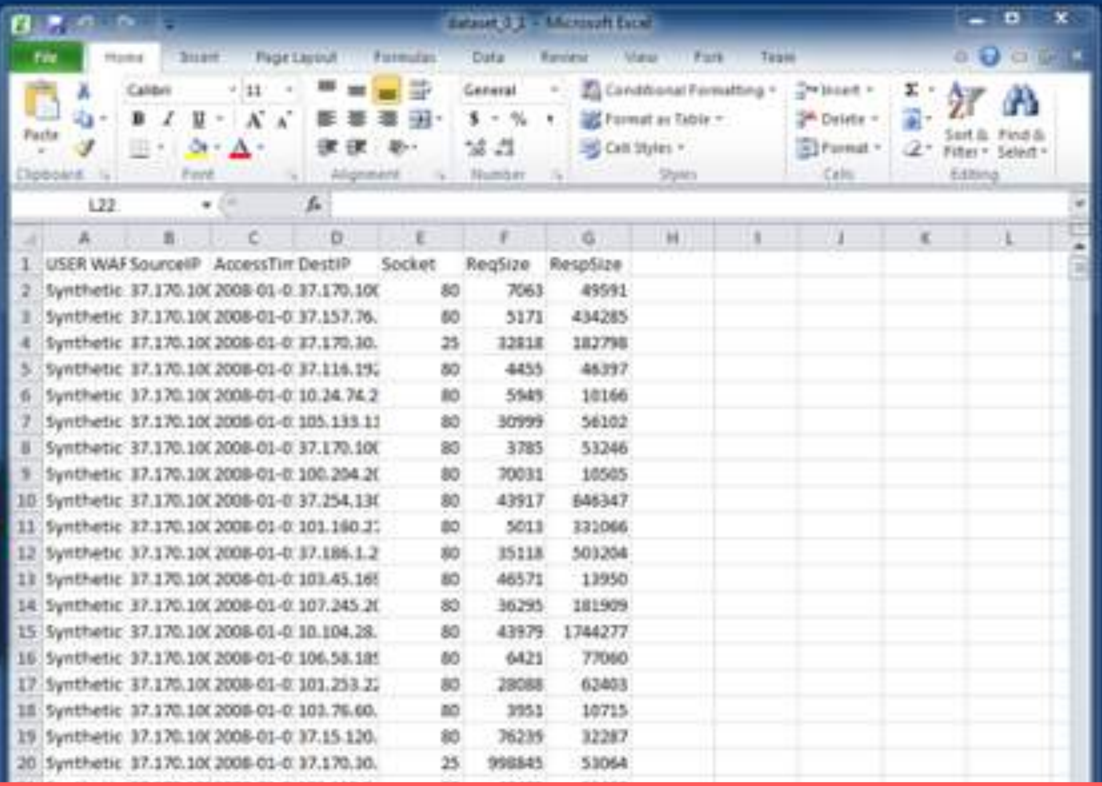
Process
Traceability



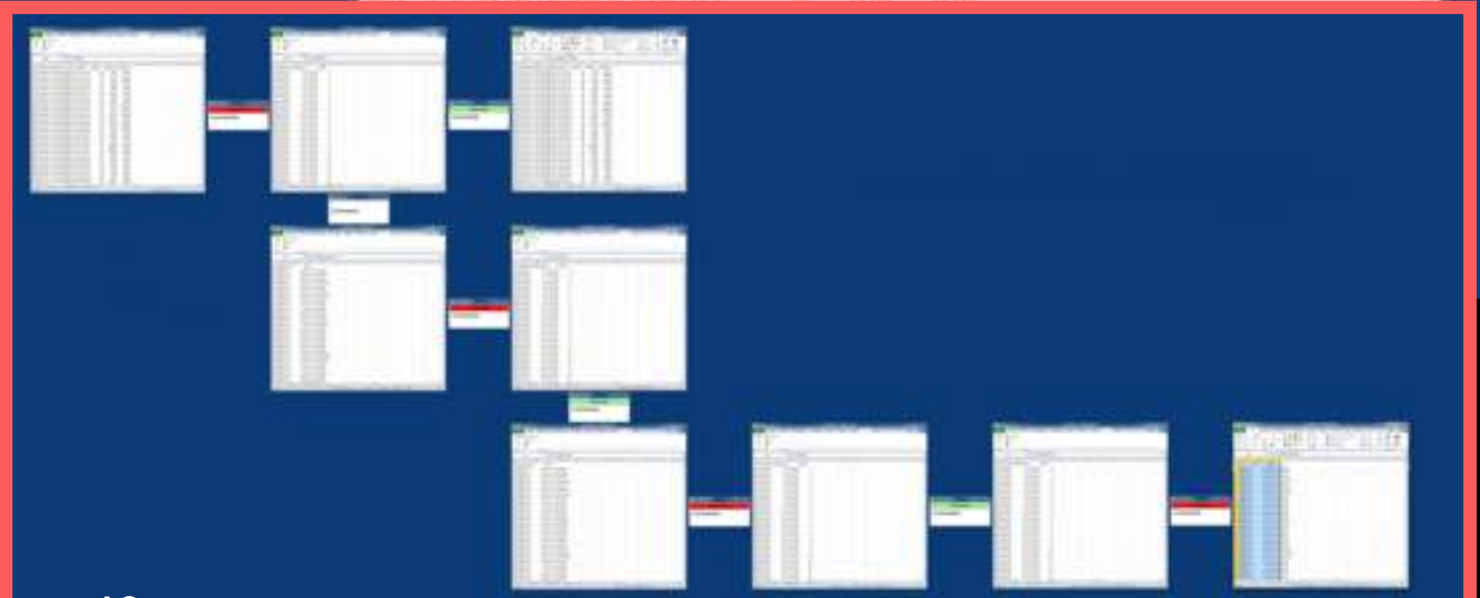
	A	B	C	D	E	F	G
1	USER WAF	SourceIP	AccessTim	DestIP	Socket	ReqSize	RespSize
2	Synthetic	37.170.10(	2008-01-0	37.170.10(	80	7063	49591
3	Synthetic	37.170.10(	2008-01-0	37.157.76.	80	5171	434285
4	Synthetic	37.170.10(	2008-01-0	37.170.30.	25	32818	182798
5	Synthetic	37.170.10(	2008-01-0	37.116.19;	80	4455	46397
6	Synthetic	37.170.10(	2008-01-0	10.24.74.2	80	5949	10166
7	Synthetic	37.170.10(	2008-01-0	105.133.11	80	30999	56102
8	Synthetic	37.170.10(	2008-01-0	37.170.10(	80	3785	53246
9	Synthetic	37.170.10(	2008-01-0	100.204.2(	80	70031	10505
10	Synthetic	37.170.10(	2008-01-0	37.254.13(	80	43917	846347
11	Synthetic	37.170.10(	2008-01-0	101.160.2:	80	5013	331066
12	Synthetic	37.170.10(	2008-01-0	37.186.1.2	80	35118	503204
13	Synthetic	37.170.10(	2008-01-0	103.45.16!	80	46571	13950
14	Synthetic	37.170.10(	2008-01-0	107.245.2(	80	36295	181909
15	Synthetic	37.170.10(	2008-01-0	10.104.28.	80	43979	1744277
16	Synthetic	37.170.10(	2008-01-0	106.58.18!	80	6421	77060
17	Synthetic	37.170.10(	2008-01-0	101.253.2:	80	28088	62403
18	Synthetic	37.170.10(	2008-01-0	103.76.60.	80	3951	10715
19	Synthetic	37.170.10(	2008-01-0	37.15.120.	80	76239	32287
20	Synthetic	37.170.10(	2008-01-0	37.170.30.	25	998845	53064
21	Synthetic	37.170.10(	2008-01-0	101.115.3.	80	48485	53420
22	Synthetic	37.170.10(	2008-01-0	107.165.1:	80	51760	55733
23	Synthetic	37.170.10(	2008-01-0	105.160.4!	80	10647	43229
24	Synthetic	37.170.10(	2008-01-0	10.133.19!	80	10760	24812
25	Synthetic	37.170.10(	2008-01-0	104.233.6!	80	4940	24911
26	Synthetic	37.170.10(	2008-01-0	101.56.23.	80	4330	33124
27	Synthetic	37.170.10(	2008-01-0	102.109.2(	80	6837	13402
28	Synthetic	37.170.10(	2008-01-0	105.141.2(	80	43107	23036



Use based comments



	A	B	C	D	E	F	G
1	USER WAF	SourceIP	AccessTim	DestIP	Socket	ReqSize	RespSize
2	Synthetic	37.170.10(	2008-01-0	37.170.10(	80	7063	49591
3	Synthetic	37.170.10(	2008-01-0	37.157.76.	80	5171	434285
4	Synthetic	37.170.10(	2008-01-0	37.170.30.	25	32818	182798
5	Synthetic	37.170.10(	2008-01-0	37.116.19;	80	4455	46397
6	Synthetic	37.170.10(	2008-01-0	10.24.74.2	80	5949	10166
7	Synthetic	37.170.10(	2008-01-0	105.133.11	80	30999	56102
8	Synthetic	37.170.10(	2008-01-0	37.170.10(	80	3785	53246
9	Synthetic	37.170.10(	2008-01-0	100.204.2(	80	70031	10505
10	Synthetic	37.170.10(	2008-01-0	37.254.13(	80	43917	846347
11	Synthetic	37.170.10(	2008-01-0	101.160.2:	80	5013	331066
12	Synthetic	37.170.10(	2008-01-0	37.186.1.2	80	35118	503204
13	Synthetic	37.170.10(	2008-01-0	103.45.16!	80	46571	13950
14	Synthetic	37.170.10(	2008-01-0	107.245.2(	80	36295	181909
15	Synthetic	37.170.10(	2008-01-0	10.104.28.	80	43979	1744277
16	Synthetic	37.170.10(	2008-01-0	106.58.18!	80	6421	77060
17	Synthetic	37.170.10(	2008-01-0	101.253.2:	80	28088	62403
18	Synthetic	37.170.10(	2008-01-0	103.76.60.	80	3951	10715
19	Synthetic	37.170.10(	2008-01-0	37.15.120.	80	76239	32287
20	Synthetic	37.170.10(	2008-01-0	37.170.30.	25	998845	53064

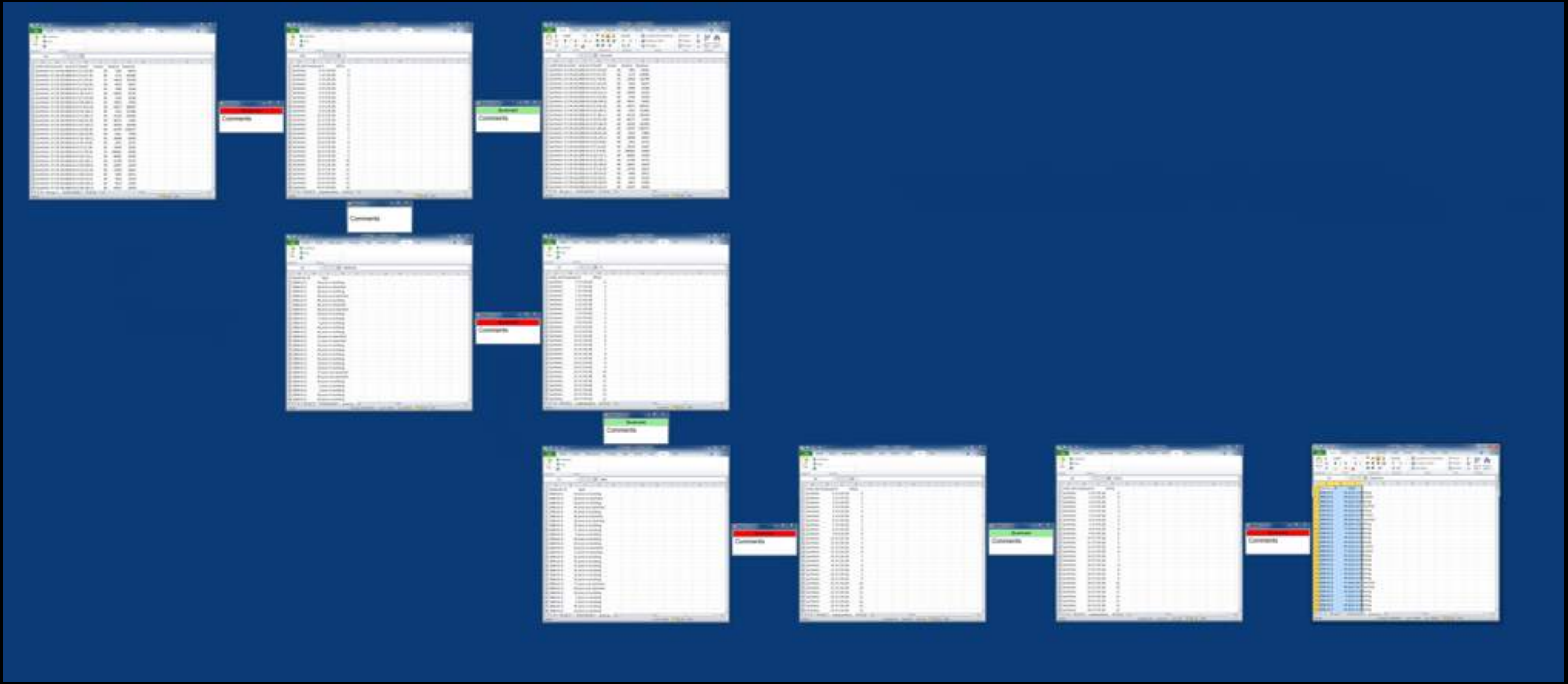


Visual History: Implementation

Branching

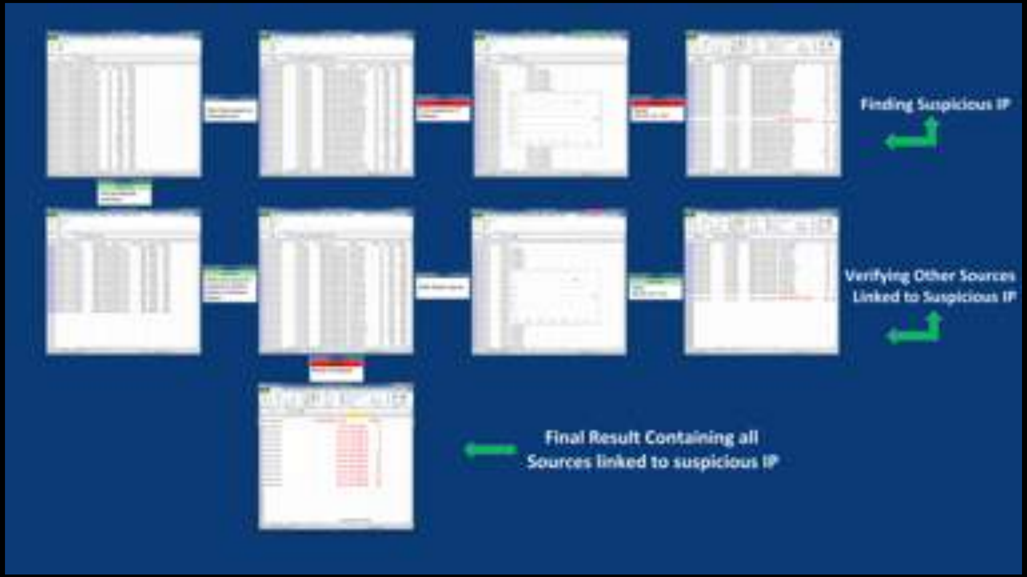
Multiple Windows, File Versions

Process Traceability



Visual History: Use Case

- 2009 VAST Challenge Dataset
  - Simulated Network Flows and Employee Building Access logs


Finding Suspicious IP
Verifying Other Sources Linked to Suspicious IP
Final Result Containing all Sources linked to suspicious IP

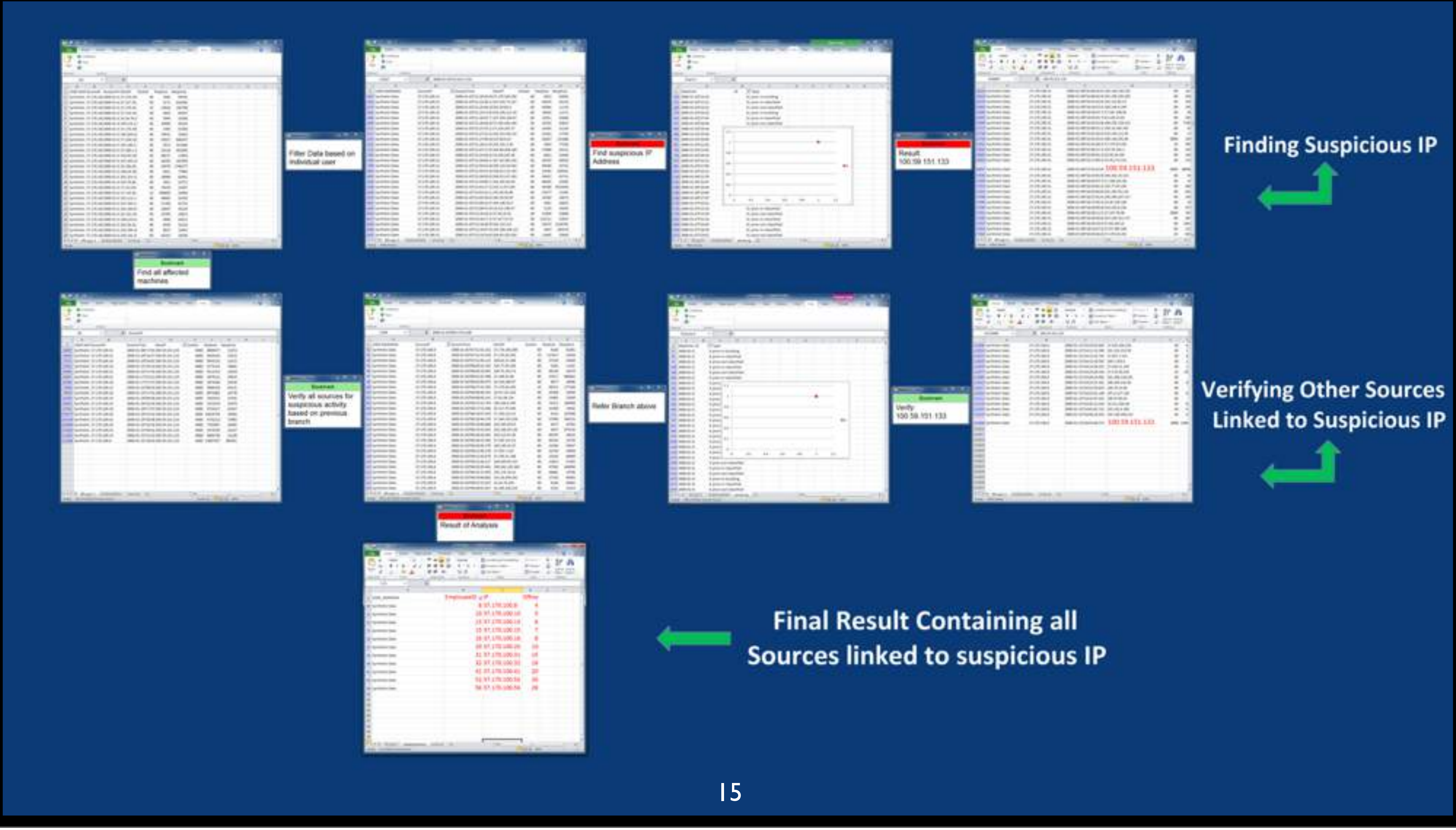
- Explore features in realistic scenario

Visual History: Implementation

Branching

Multiple Windows,
File Versions

Process Traceability



Bookmark
Filter Data based on individual user
Bookmark
Find suspicious IP Address
Bookmark
Result 100.59.151.133
Bookmark
Find all affected machines
Bookmark
Verify all sources for suspicious activity based on previous branch
Refer Branch above
Bookmark
Verify 100.59.151.133
Bookmark
Result of Analysis

Finding Suspicious IP

Verifying Other Sources Linked to Suspicious IP

Final Result Containing all Sources linked to suspicious IP

Use Case: Lessons Learned

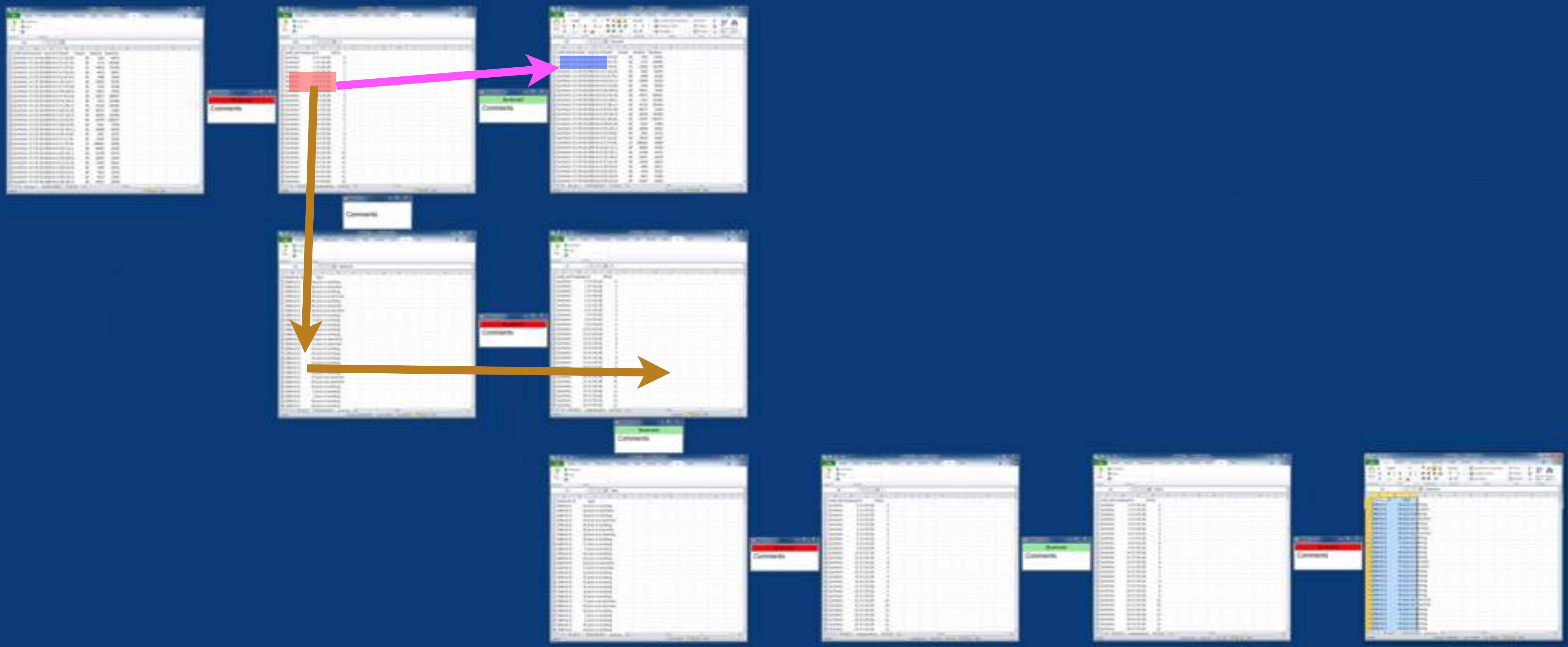
- Propagating Changes vs. Branching new version
  - Brushing & Linking through Process

- Automatic vs. Manual Layout of History
  - When to fork, branch?
  - Running out of space?

Propagating vs. Forking

Visual History maintains process actively in the workspace.

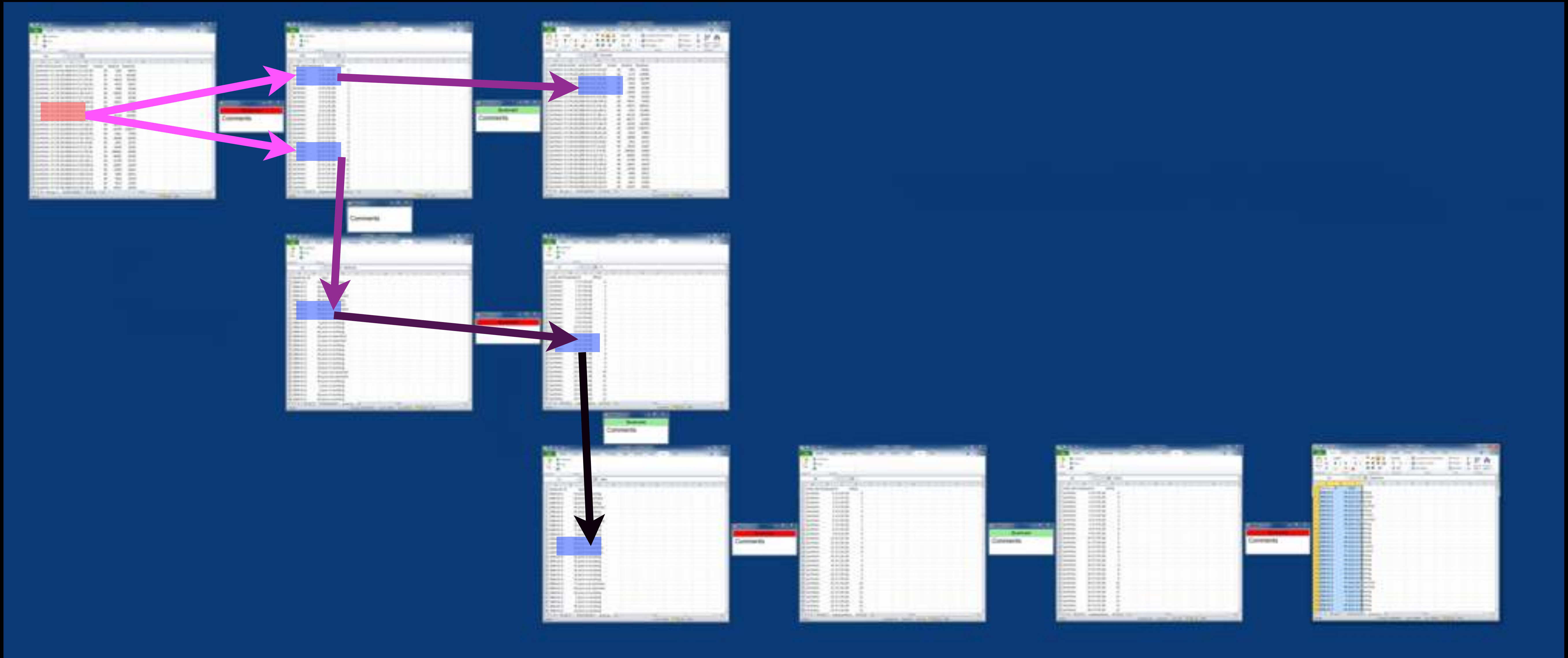
How to adjust workspace when previous states are changed?



Brushing & Linking through Process

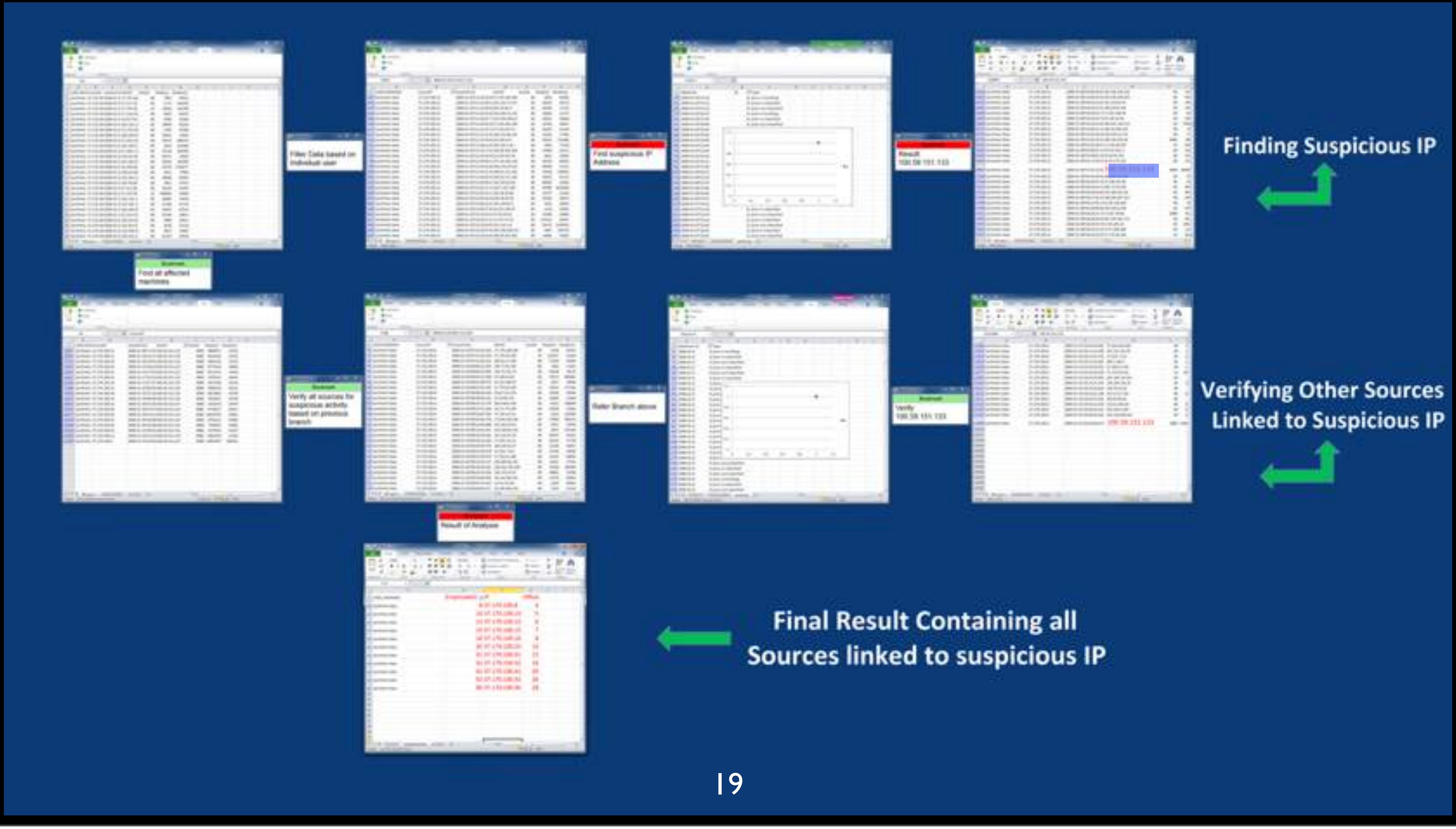
Visual History maintains process actively in the workspace.

How to highlight impacted downstream data?



Automatic vs. Manual Layout

- *Balancing automatic branching with user-defined positioning of windows*
- *How to handle display space limitations?*
- *Scalability of branching*


Filter Data based on individual user
Find suspicious IP Address
Result 100.58.151.133
Finding Suspicious IP
Find all affected machines
Verify all sources for suspicious activity based on previous branch
Refer Branch above
Verify 100.58.151.133
Verifying Other Sources Linked to Suspicious IP
Result of Analysis
Final Result Containing all Sources linked to suspicious IP

Future Work

- Evaluate design decisions from lessons learned
  - Implementation
- Formal user study evaluation
  - *How does keeping the history current impact the dynamic analytic process of the user?*

Conclusions

- Cyber Analytic Workspaces can support the *process* of the analyst
  - Combining algorithmic aids (e.g., sniffers, filters, alerts, ...) with human intuition
- With *Visual History*, we merge traditional "history" with "process"
- *Visual History* focuses on the importance of the user *process* as well as the *solution*

Conclusions

- Cyber Analytic Workspaces can support the *process* of the analyst
    - Combining algorithmic aids (e.g., sniffers, filters, alerts, ...) with human intuition
- With *Visual History*, we merge traditional "history" with "process"
- *Visual History* focuses on the importance of the user *process* as well as the *solution*

Thanks! Questions?