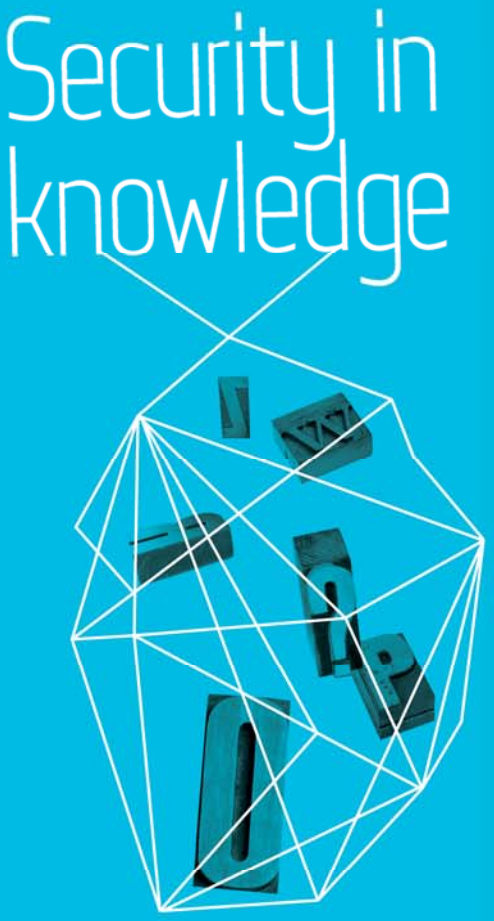
RSA CONFERENCE ASIA PACIFIC 2013



ACCELERATING THE ANALYST WORKFLOW: LEARNING FROM INVESTIGATIVE ACTIONS

Dennis Moreau
RSA, The Security Division of EMC

Samir Saklikar
RSA, The Security Division of EMC

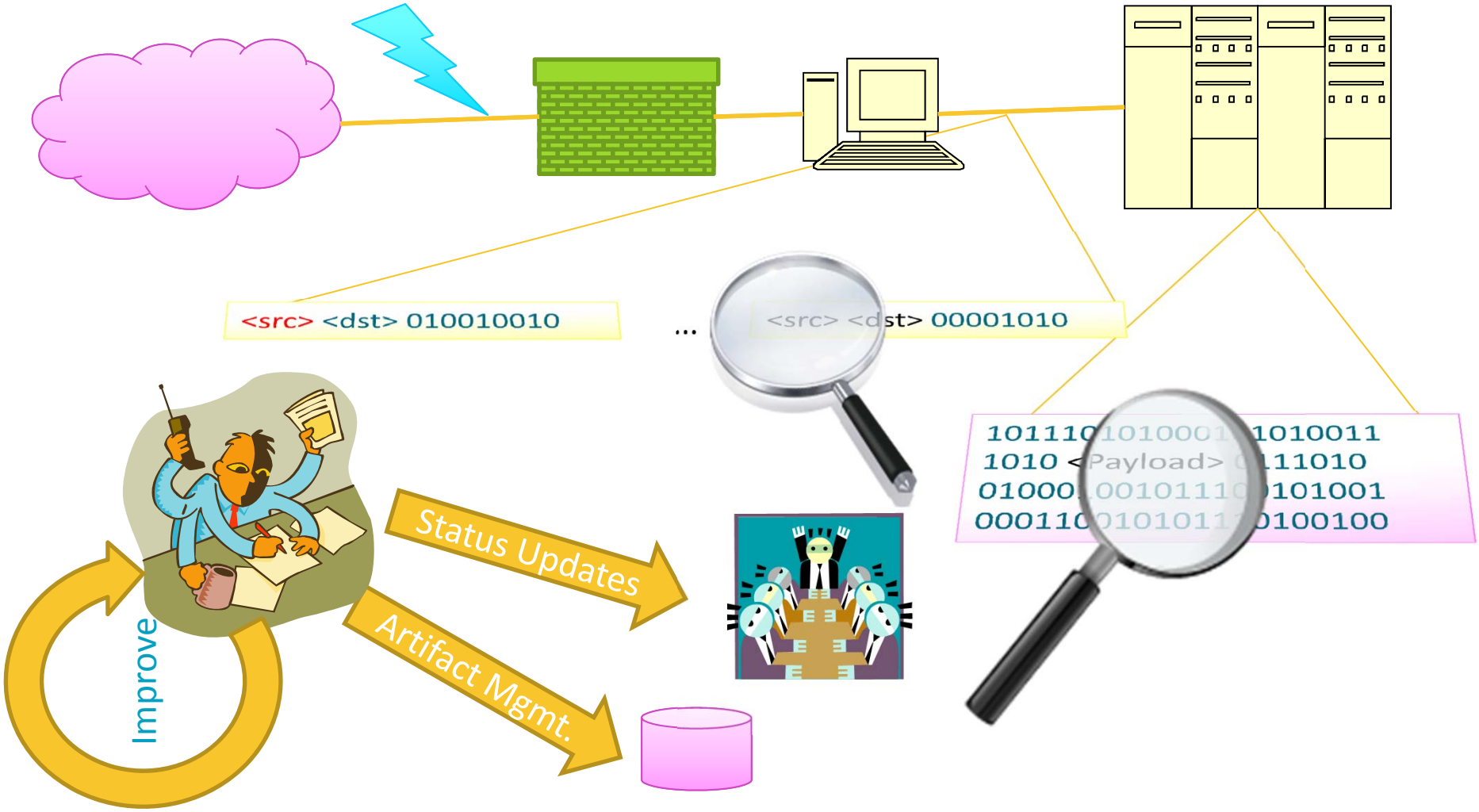
Session ID: CLE-W01

Session Classification: Intermediate

Agenda

- Motivation
- Requirements on the tool
- A PoC implementation
- Conclusions and Directions

The Problem


<src> <dst> 010010010
...
<src> <dst> 00001010
Status Updates
Artifact Mgmt.
Improve

Motivation – Improve Analytic Workflows

- Inefficiencies
  - Investigation/response status updates are disruptive
  - Manual management of forensic artifacts is error prone
  - Analytic/investigative process documentation is tedious
  - Retrospective recall is incomplete, limiting improvement and team/organizational learning
- Opportunities
  - Demonstrated behavior as a basis for ongoing trust – as sharing expands
  - Documented response as part of response service delivery
  - Community analytic and investigative pattern learning

Approach – Analyst Activity Charts

- Discovery
  - Monitor and track all actions done by analyst while handling an incident
- Documentation
  - Document into data-sequential flow, tracking dead-ends and iterations
- Attestation
  - Time-based cryptographic hashes for attestable proof of records
- Automation
  - Convert charts into workflows, and automate as much possible

Video Demo

- "Red October" [targeting MS Excel/Word exploits for cyber-espionage]
- Analysis on www.malware.lu

What data should be recorded?

- Tools
  - SIEM events, Packet Capture, Sandboxes, Reversing tools, custom scripts, ~ to notepad.exe
- Input artifacts
  - Log data, Packet stream, binaries, clipboard buffers, memory dumps, registry
- Analyst Inputs
  - Key-presses, commands, mouse-clicks
  - Network and File system activity from workstation
- Intermediate data files
  - IDB files, extracted packets, binaries
- Output artifacts/Results
  - File-based output, Screen-based output, Human-derived outputs

Recording Inputs and Analyst commands

- Recording Input artifacts
  - Can link into existing incident/case management system
  - Can also be tracked separately
- Collecting Analyst commands/inputs
  - Windows - Tap into the Command prompt shell
    - Enable Telnet into windows system; Enable logging in putty
    - Enable time-stamps in the prompt ($P:$T$G)
  - Linux –

```
export PROMPT_COMMAND='RETRN_VAL=$?;logger -p local6.debug
"#$(history 1 | sed "s/^[ ]*[0-9]\+[ ]*//" )"‘

vi /etc/syslog-ng/syslog
[…]
destination d_usercommands{ file("/var/log/usercommands.log"); }
```

Monitor the file-system, clipboard, graphical applications

- Python (watchdog package) or inotify-tools on linux

```
watchmedo log –patterns =“*.log” –ignore-directories –
recursive .
inotifywait -qmr --timefmt "%X" -o ~/inotifywait.log --format
"%T#%w#%f#%e“ .
```

- Python-based clipboard monitor for text copies
  - Pyperclip - http://coffeeghost.net/src/pyperclip.py
  - Simple polling works

- Tracking currently selected graphical applications

```
xprop -id $(xprop -root 32x '\t$0' _NET_ACTIVE_WINDOW | cut -f
2) '\t$0' _OB_APP_NAME | cut -f 2
```

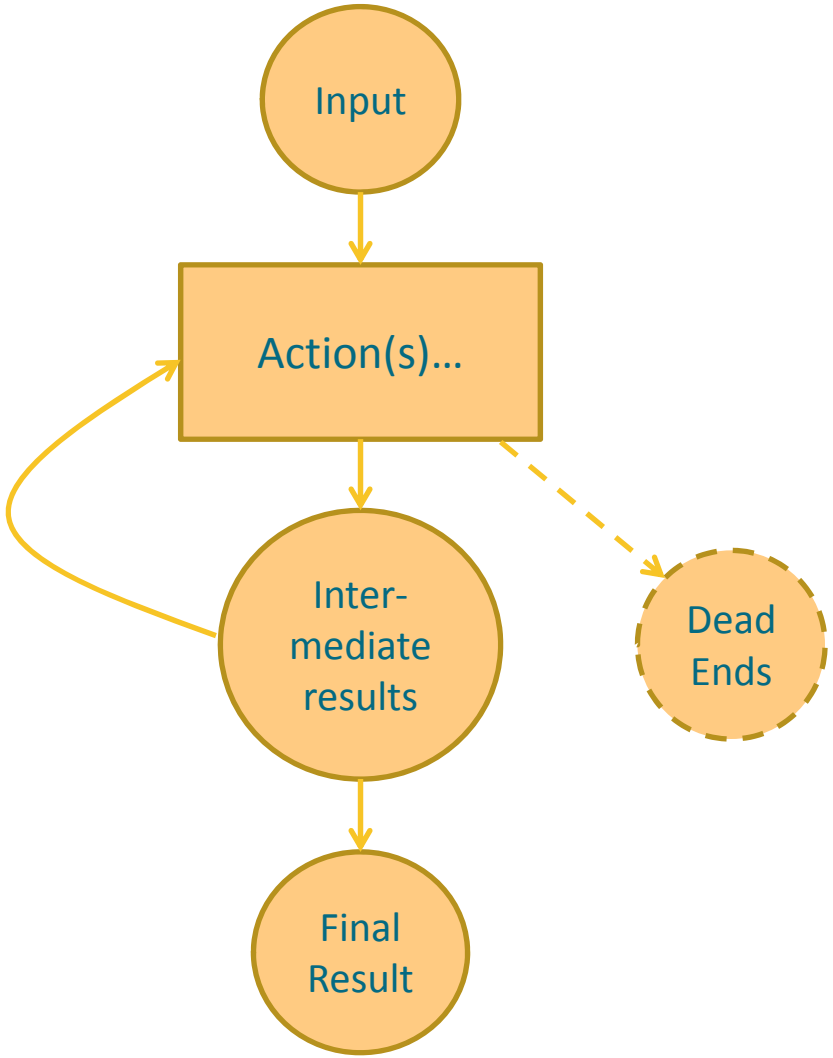
Monitor network and reversing activity

- Everyone has a favorite here
- Wireshark, Netmon, Bro etc. to log and extract files
- Fiddler2 - http://www.fiddler2.com/fiddler2/
  - Filters to remove irrelevant domains (google-analytics.com, youtube.com)
  - Export files and session data (json format)
  - Parse to extract requests and uploads and responses.
- Netwitness Investigator freeware

- IDA Pro CollabREate Plugin
  - Multiple IDA instances sync with a CollabREate server
  - Dynamic updates are pushed since last user session
  - Track IDB changes

Documentation (normalized representation)

- Chart
  - <Input-data, analyst-actions, output-results, time>…
- <Input-data>
  - <log-data, network-data, endpoint-data, binary-data, prior-results,…>
- <analyst-actions>
  - <tools-used, services-invoked, key-presses, mouse-clicks, commands,…>
- <output-results>
  - <modified-network-data, modified-endpoint-data, modified-binary-data, results, final-resolution,…>



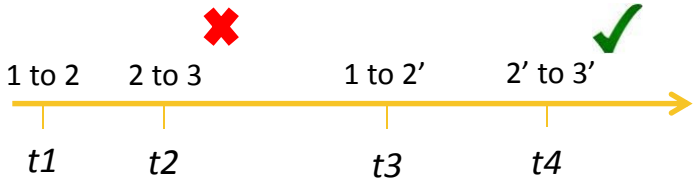
Normalizing the Data

- Pre-Processing the Data
  - Time-stamps of all User Actions, File Access, GUI applications
  - Unique Identifiers of Files, data objects by hashing the content/filename
  - Unique Identifiers for actions (commands, UI interactions) based on time
    - Using "vi" at different times will result in different action nodes.
    - Avoid commonly used tools to be become high-edge single nodes.
    - Different invocations of a tool may have been used for different purposes (with different parameters/options)

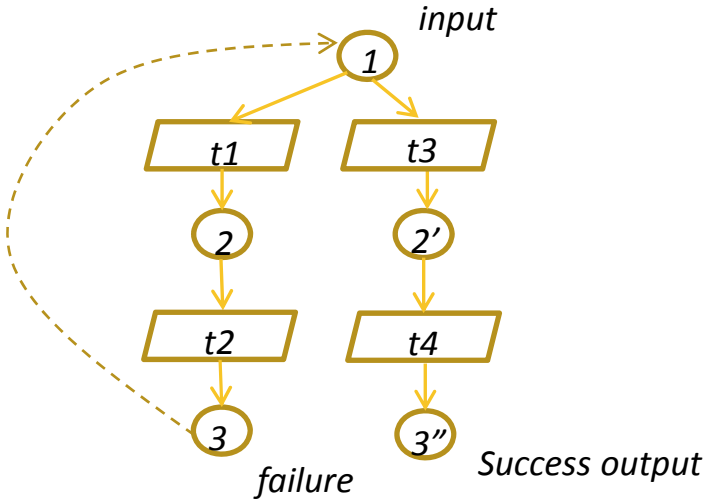
- Normalize to a common format
  - Nodes in a graph structure, represent both actions and artifacts
  - Node = {id, label, time, command/artifact uri/content}

Correlating Analyst Actions/Inputs and Outputs

- How to go from time-sequential actions to data flows?
- Adding Edges for data transformations
  - Time correlate actions and artifacts (same or close in time-stamps)
  - Track file actions (CREATE, MODIFY, DELETE, CLOSE_NOWRITE, CLOSE_WRITE) to determine input files and output files.

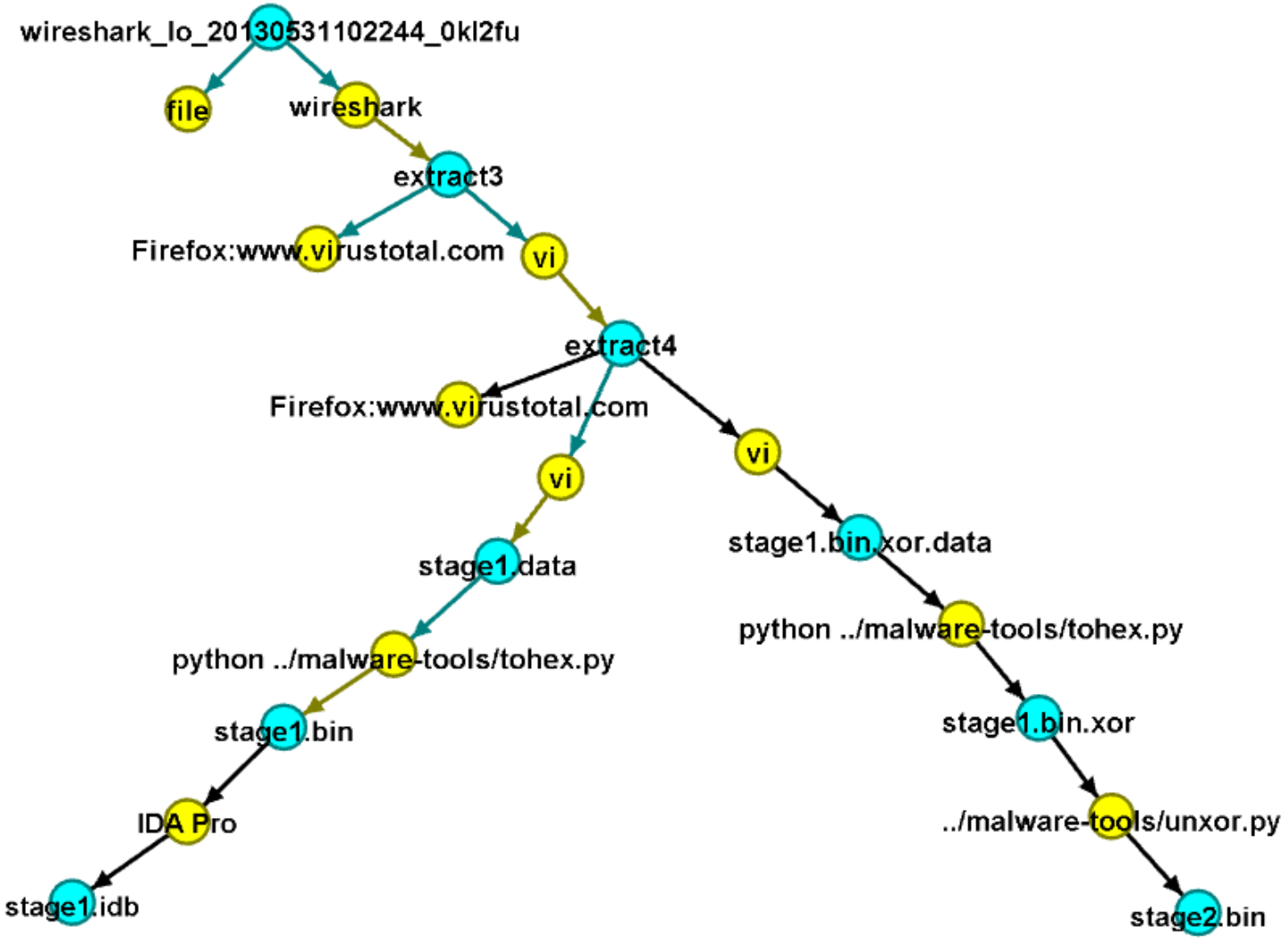


Analyst Actions on input 1 to reach output 3'

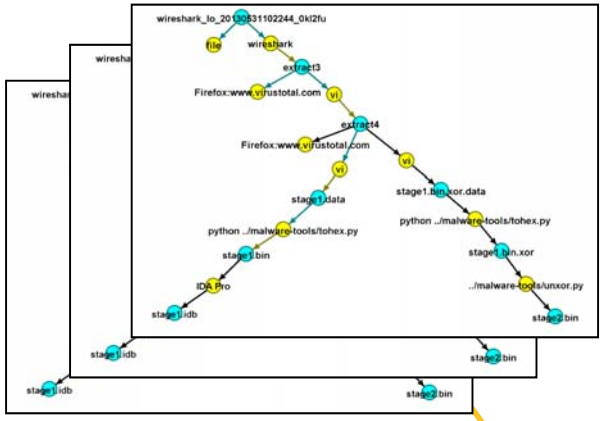


- Identify and subsume "copy-paste" actions (which could result in broken links)
- Normalize directory structure/naming to correlate across windows/linux workstations

Analyst Activity Chart (data-flow representation)



Attestation (Auditing and Intelligence Sharing)

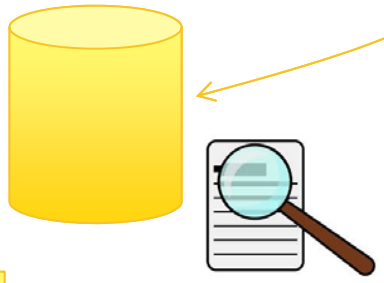


hash

Real-time streaming

3rd party Compliance/Legal assessment.

Discrete Incident-based Auditing



Long-term archival & searching

Privacy & Compliance filters

Augmenting Threat Intelligence feeds

Conclusions

- Improved investigative process visibility supports:
  - Improved efficiency
    - Status Visibility and Update Automation – Where are we in the response? What do we know? Who is doing what?
    - Forensic Artifact Management – What tools ingested which data and produced which files?
  - Enhanced experiential Learning
    - What did we do? (process documentation)
    - Based on comparison across time or across analysts, what should we Standardize? Automate? Never do again?
    - How can we improve? Cooperate? Collaborate?
  - Enhanced Trust
    - How did we handle indicators/forensics?
    - Is our performance and practice a basis for extended trust?
    - Did we demonstrate due care?

Next Steps: Responder Activity

- Instrument responder environment for capturing ECAT usage by responders.
- Analyze captured activity across the response cycle (Tier 1-Tier 3)
- Many aspects of response are more operationally regular than are those of reversers and MW analysts, so:
  - Determine potential KPIs
  - Determine completeness of responder behavior capture
  - Potentially extend instrumentation or logging for more coverage
  - Construct a dimensional schema to support analysis of responder efficiency and efficacy

Questions