# An Empirical Study of Automatic Event Reconstruction Systems

*By*

**Sundararaman Jeyaraman, Mikhail Atallah**

*Presented At*

The Digital Forensic Research Conference

**DFRWS 2006 USA** Lafayette, IN (Aug 14th - 16th)

# An Empirical Study of Automatic Event Reconstruction Systems

Sundar Jeyaraman
Mikhail J Atallah

## Event Reconstruction

- Identify the underlying conditions and chain of events that led to the security event

- Necessary for effective incident response and recovery

PURDUE UNIVERSITY

## Event Reconstruction cont.

- Ex-post evidence
  - Disk, Memory dumps. Network logs
  - TCT, Sleuthkit, Encase, Ethereal…
- Ex-ante logging
  - Audit trails (hopefully tamper proof)
  - Back Tracker, Forensix…

## Why an empirical study?

- Guidance for investigators in choosing the right tool
- Likelihood calculation for hypotheses
- Towards standardization and thwarting Trojan Horse Defense [Carney et al. 2004, ]
- Directions for future research

PURDUE UNIVERSITY

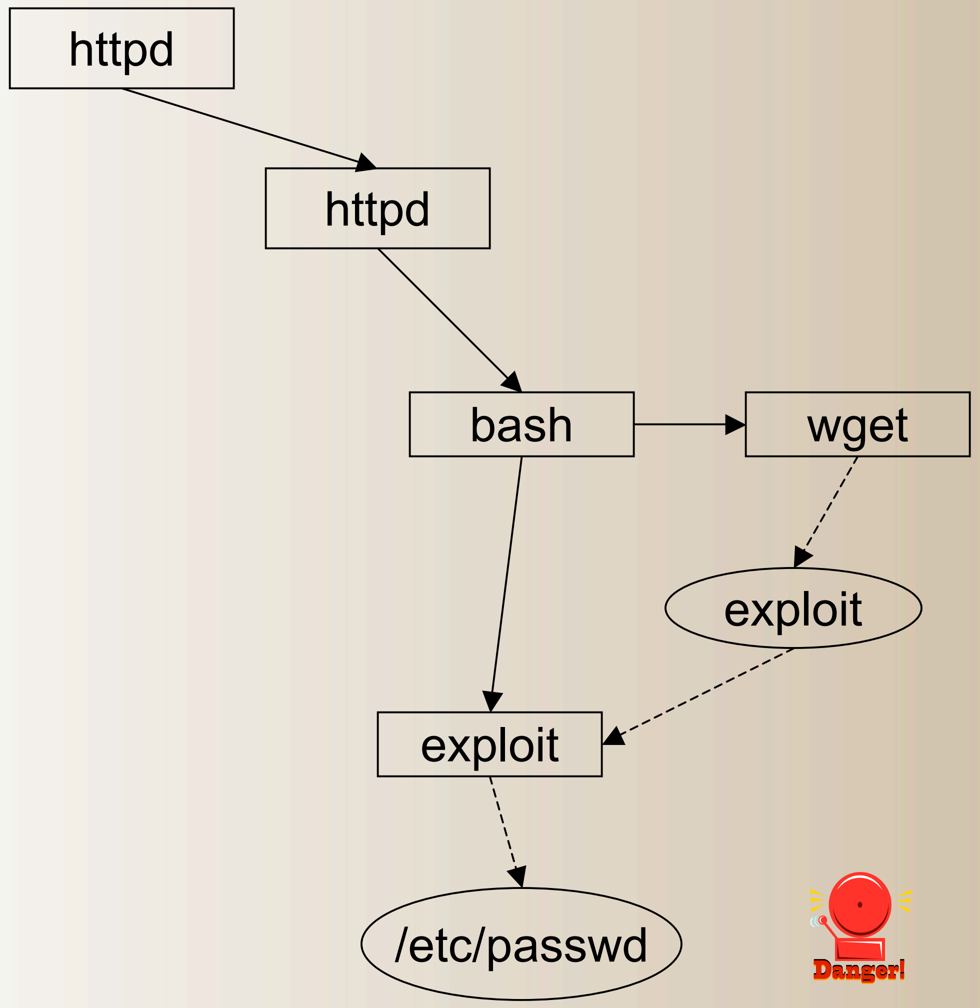
## A really quick survey of event reconstruction systems

CERIAS

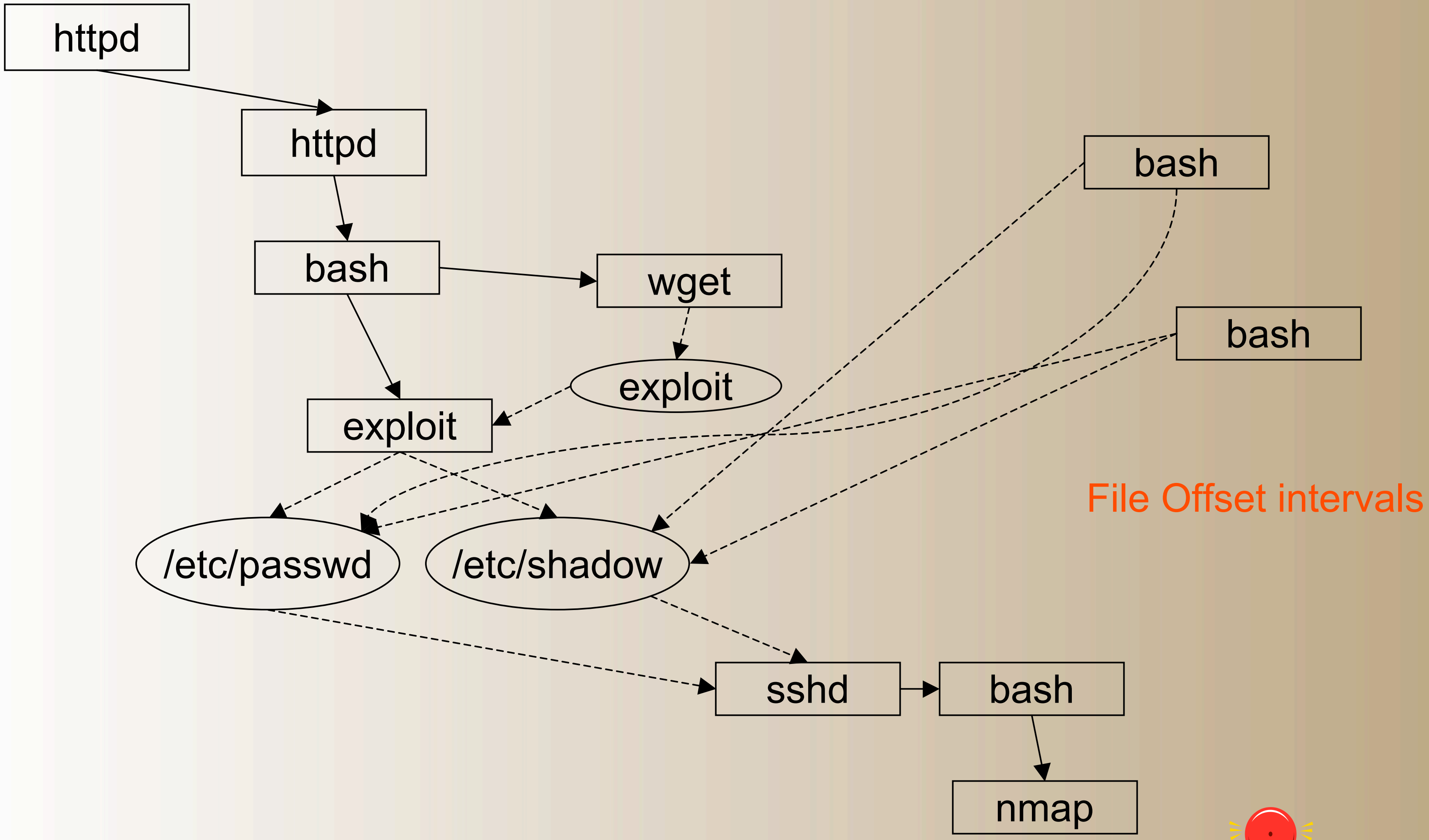
DFRWS 2006

## BackTracker [king et al. 2003]

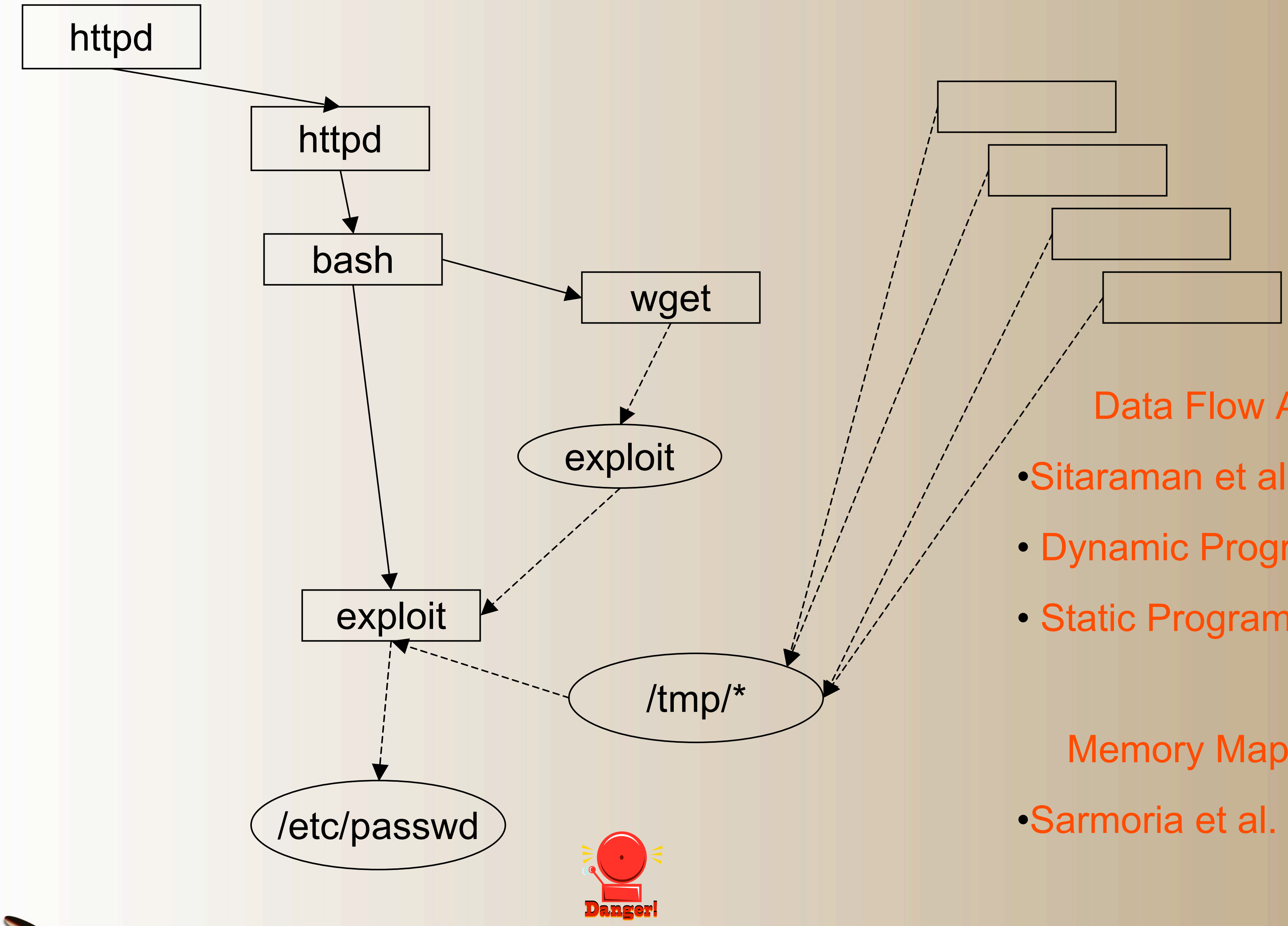
- At run time:
  - Monitor system objects and events
  - Record dependences between system objects
- Post-mortem:
  - Build dependence graph
  - Traverse graph to reconstruct the events


httpd
httpd
bash
wget
exploit
exploit
/etc/passwd

Danger!


httpd
httpd
bash
wget
exploit
exploit
bash
bash
/etc/passwd
/etc/shadow
sshd
bash
nmap
File Offset intervals

Danger!



Data Flow Analysis

- Sitaraman et al. 2005
- Dynamic Program Slicing
- Static Program Slicing

Memory Mapped Files

- Sarmoria et al. 2005

## In Summary

- Tracking OS-enabled dependences
  - BackTracker, Forensix, CIDS, Process Labels
  - Improved BackTracker
    - File Offset Intervals
- Tracking process-enabled dependences
  - Improved BackTracker
    - Static Program Slicing
    - Dynamic Program Slicing
  - Memory mapped files

## Methodology

- Equivalent ability in tracking causal relationships enabled by the OS.
- Difference arises in the ability to track those enabled by the process address space
- Use *dynamic slicing* to determine false-positives and false-negatives

# Reconstruction Systems

- Tracking OS-enabled dependences
  - **BackTracker, Forensix, CIDS, Process Labels**
  - Improved BackTracker
    - File Offset Intervals

- Tracking process-enabled dependences
  - **Improved BackTracker**
    - **Static Program Slicing**
    - **Dynamic Program Slicing**
  - Memory mapped files

## Methodology cont.

- A set of applications as a benchmark suite
- Regression test suite for each application
- Metrics
  - Average rate of false-positives

## BenchMark Suite

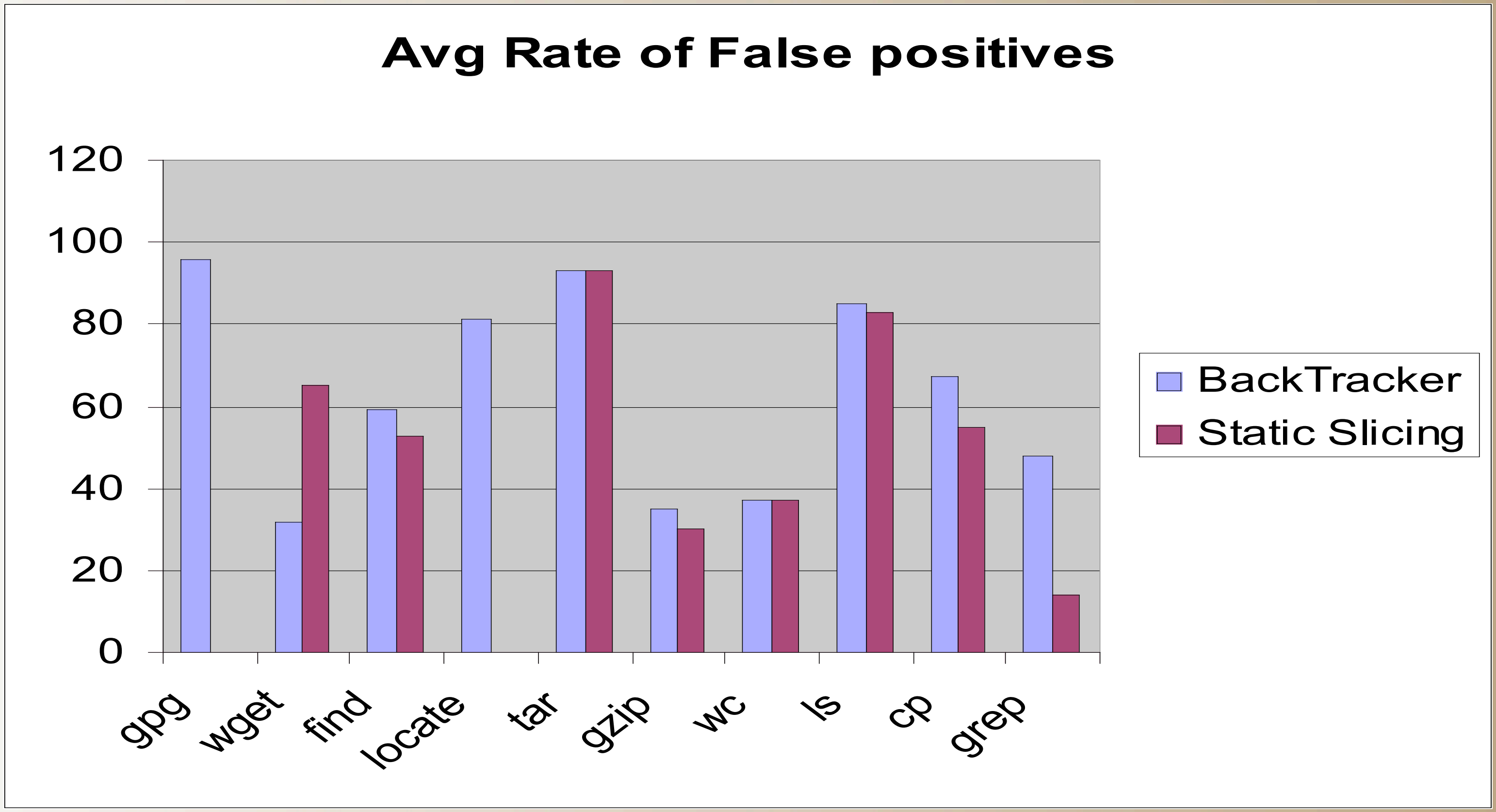
| | |
|---|---|
| gnuPG 1.4.2 | GNU's replacement for PGP |
| gnu wget 1.10 | Program for retrieving files through HTTP(S), FTP |
| find (findutils 4.2.25) | Search for files in a directory hierarchy |
| locate (findutils 4.2.25) | List files in a database that matches pattern |
| ls (coreutils 4.5.3) | List directory contents |
| cp (coreutils 4.5.3) | Copy files |
| wc (coreutils(4.5.3) | Print the number of bytes, words and lines in a file |
| tar 1.15.1 | Archiving software |
| gzip 1.3.3 | A popular data compression program |
| grep 2.5.1 | Search files for a given input pattern |

## Experimentation

- Dynamic slicing implemented using PIN
- Static Slicing implemented using CodeSurfer
- Approx. 100,000 system calls and 11 Billion instructions executed as part of the test cases

PURDUE UNIVERSITY

## Avg rate of False-positives



CERIAS

## Overhead of Dynamic Slicing

| Application | CPU overhead | Wallclock overhead |
|---|---|---|
| gpg | 8458 | 7646 |
| wget | 4933 | 45 |
| find | 648 | 648 |
| locate | 43298 | 48 |
| tar | 12808 | 14149 |
| gzip | 32894 | 1510 |
| wc | 28719 | 760 |
| ls | 22153 | 8140 |
| cp | 10502 | 11525 |
| grep | 53 | **57** |

## Limitations & Discussion

- Incomplete coverage of reconstruction systems
- Limitations of benchmark suite
  - No multi-threaded applications
  - No application > 100K LOC
- No statement coverage statistics for testcases
- Implicit dependences
- Better analysis of the results

Comments/Questions/Brickbats?

Sundar Jeyaraman

jsr@cerias.purdue.edu

# Iterative and Recursive Behavior

```
while (pending_dirs)
{

 extract_files_from_dir(pending_dirs);
 print_files();
}
```

```
  dir1     dir2  dir3
    |        |     |
--------    f2    f3
|      |
f1     d1
```

\`ls dir1 dir2 dir3\`