pervasivetechnologylabs
AT INDIANA UNIVERSITY

www.pervasivetechnologylabs.iu.edu

Honeynet Data Analysis:

A technique for correlating sebek and network data

Edward G. Balas

Indiana University

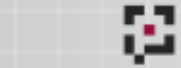
Advanced Network Management Lab

6/15/2004

pervasivetechnologylabs
AT INDIANA UNIVERSITY
www.pervasivetechnologylabs.iu.edu

About the Author

- Edward G. Balas
  - Security Researcher at Indiana University's Advanced Network Management Lab.
  - Honeynet Project Member
    - Sebek project lead
    - Honeywall User Interface project lead
- Research Sponsorship
  - This materials based on research sponsored by the Air Force Research Laboratory under agreement number F30602-02-2-0221. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.



pervasivetechnologylabs
AT INDIANA UNIVERSITY
www.pervasivetechnologylabs.iu.edu

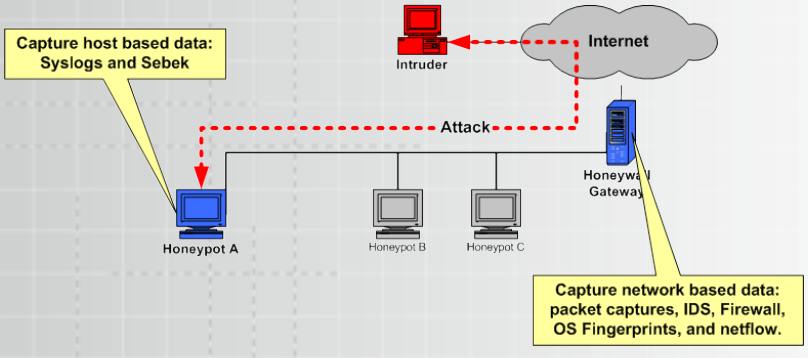
Roadmap

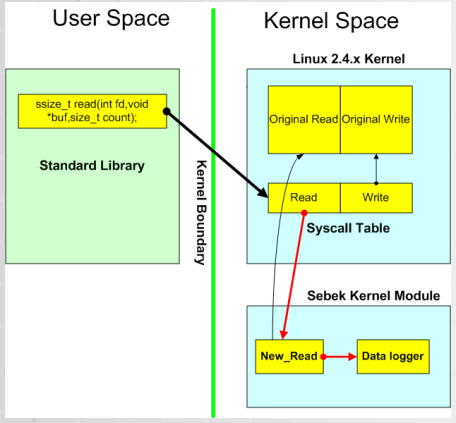
- Honeynets are an idealized forensic testbed
- These testbeds have lead to a new data capture tool called Sebek.
- The volume of data has precluded use in operational environments.
- Describe efforts to solve issue by enhancing Sebek.
- Hope to provide quicker examination of data
- May yield a viable tool for forensics.

Introduction to Sebek

- Sebek Data Capture tool
  - kernel space tool that monitors sys_read call
  - covertly exports data to server.
  - used to monitor keystrokes, recover f les, and other related activities even when session encryption used.
  - http://www.honeynet.org/tools/sebek/

Sebek Illustrations

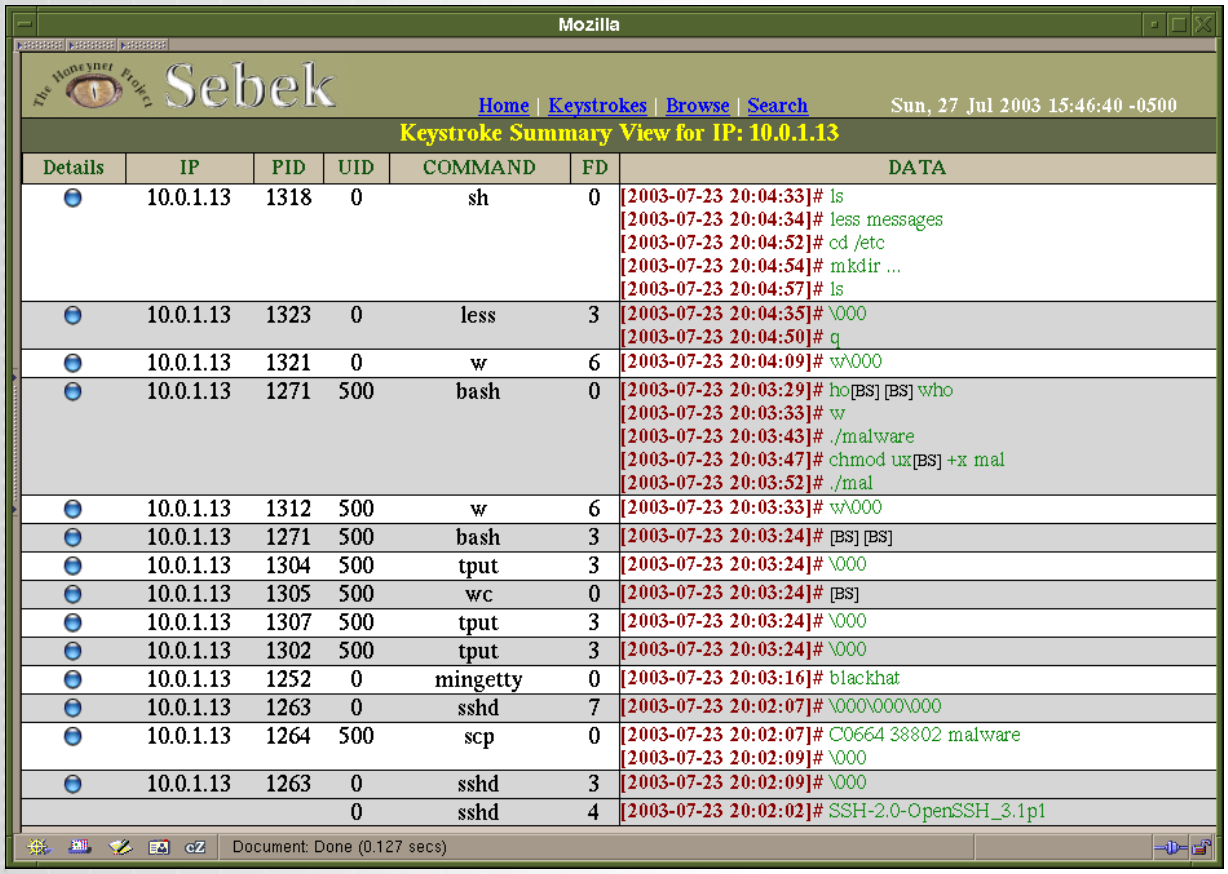




- top left shows general architecture
- bottom left provides illustration of how Sebek gains access to sys_read data.

pervasivetechnologylabs
AT INDIANA UNIVERSITY
www.pervasivetechnologylabs.iu.edu

What the data "looks" like



Existing Capabilities

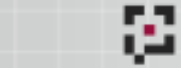
- What this gives you
  - Keystrokes
  - Files copied to system with session encryption
  - Burneye passwords
  - Read activity for each process.
- What is missing
  - Way to f ltter or navigate the volume of data
  - Sense of relationship between processes
  - Correlation to IDS or other network events.
  - Names of Files associated with File Descriptor

Enhancements to Sebek

- Record Socket Information
  - allows us to correlate network events to the associated process , user and even f le descriptor on a box running sebek.
- Record Fork and Parent PID information
  - allows us to rebuild the process tree
  - combined with Socket Info, provides a fault tree.
- Record all f les Opened
  - identify all f les “touched” in association with with an event.

Socket Monitoring

- To correlate network connections to process / f le number we added the ability to monitor the sys_socket call.
  - in Linux, all socket calls are multiplexed through one generic socket call.
  - gained access using the same technique as used with sys_read.
  - this provided a mapping of:
    - src/dst ip endpoints for a connection
    - src/dst ports and protocol
    - state of connection.
    - Related Process, File No, etc.



pervasivetechnologylabs
AT INDIANA UNIVERSITY
www.pervasivetechnologylabs.iu.edu

Parent PID tracking

- Record the process inheritance tree by reporting the Parent PID along with the PID
  - Each sys_read provides the Parent PID
  - Each sys_fork provides a mapping as well.
    - needed because not all processes read before forking.

Data Analysis

- Honeynet data analysis and the analysis of network based intrusions are quite similar.
- Multiple Data types examined
  - Network traffic logs
  - IDS / Event logs
  - Disk Analysis
  - Sebek or other keystroke logs
- Time consuming and error prone.

Three steps in analysis

- Collect/Screen
  - Identify raw data of interest
- Coalesce
  - Combine data from different data sources, identifying cross data source relations and providing some type of normalized access to the data.
- Report
  - Identify central themes, screen out superf luous data.

How it is done today

- Each data type has its own analysis tool
  - causing a stovepipe effect.
  - each data set goes through the 3 steps in isolation.
- Switching data sources causes wetware context switch.
- Relations manually discovered and expressed to each tool for screening by analyst.
- No automatic way to track interesting sequences across data sources.

Why this is no good

- Labor intensive
  - I am lazy
- Error Prone
  - I am sloppy
- Lots of menial work being done by a human
  - I paid a lot for this computer

Where we want to be

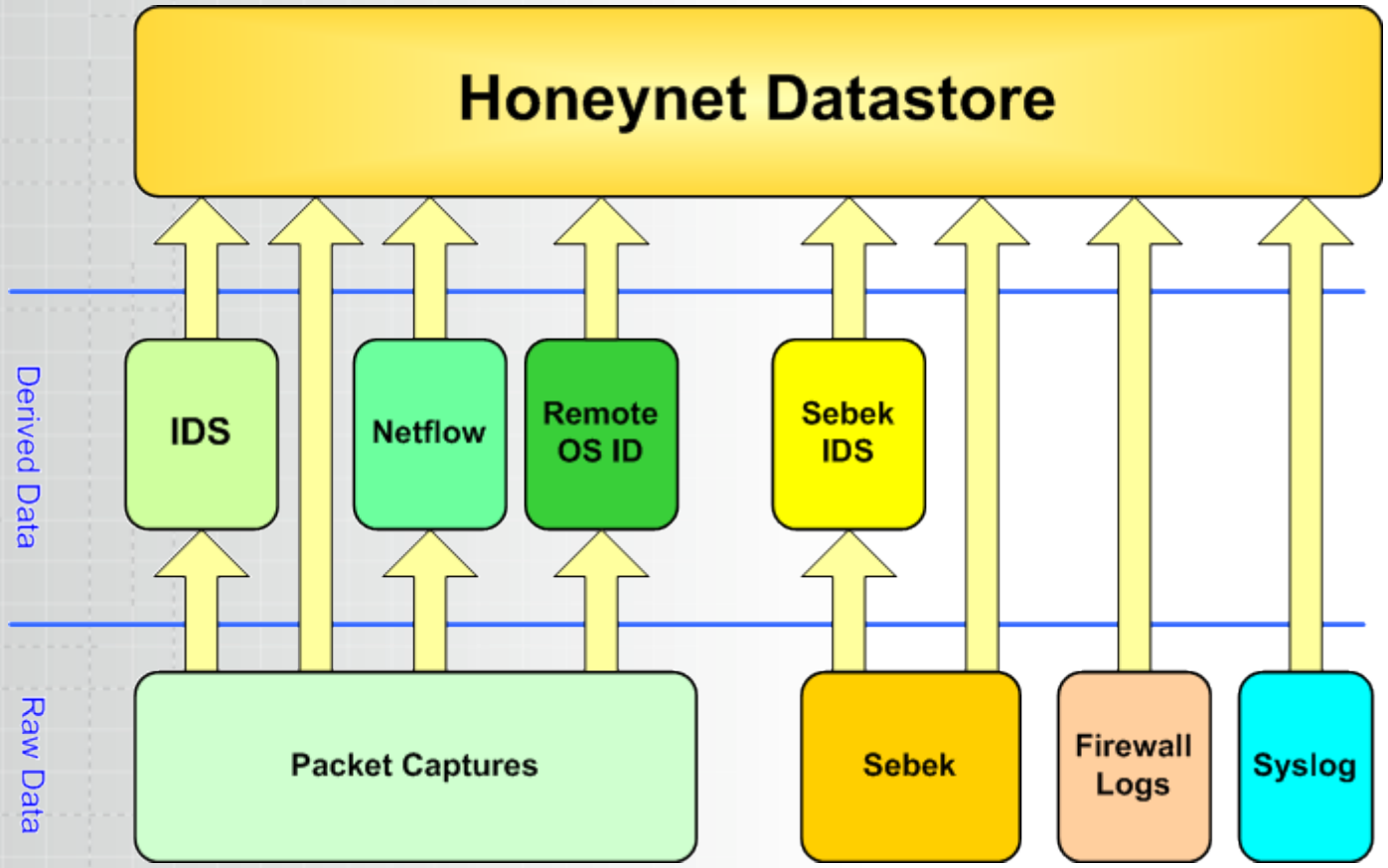
- Shift the Screening and Coalescing burden to the computer.
- Focus human effort on tasks best suited to the human.
- Provide an interface that supports the analyst's workf bw.
- Provide a system that may have use in production networks.

Improving Data Analysis

- The new data coming from sebek allows us to automatically relate network and sebek data.
- To automate coalescing we developed a backend daemon called Hf bw.
- To demonstrate the impact of these capabilities on reporting, we developed a web based user interface named Walleye.

pervasivetechnologylabs
AT INDIANA UNIVERSITY
www.pervasivetechnologylabs.iu.edu

The challenge facing Hf bw

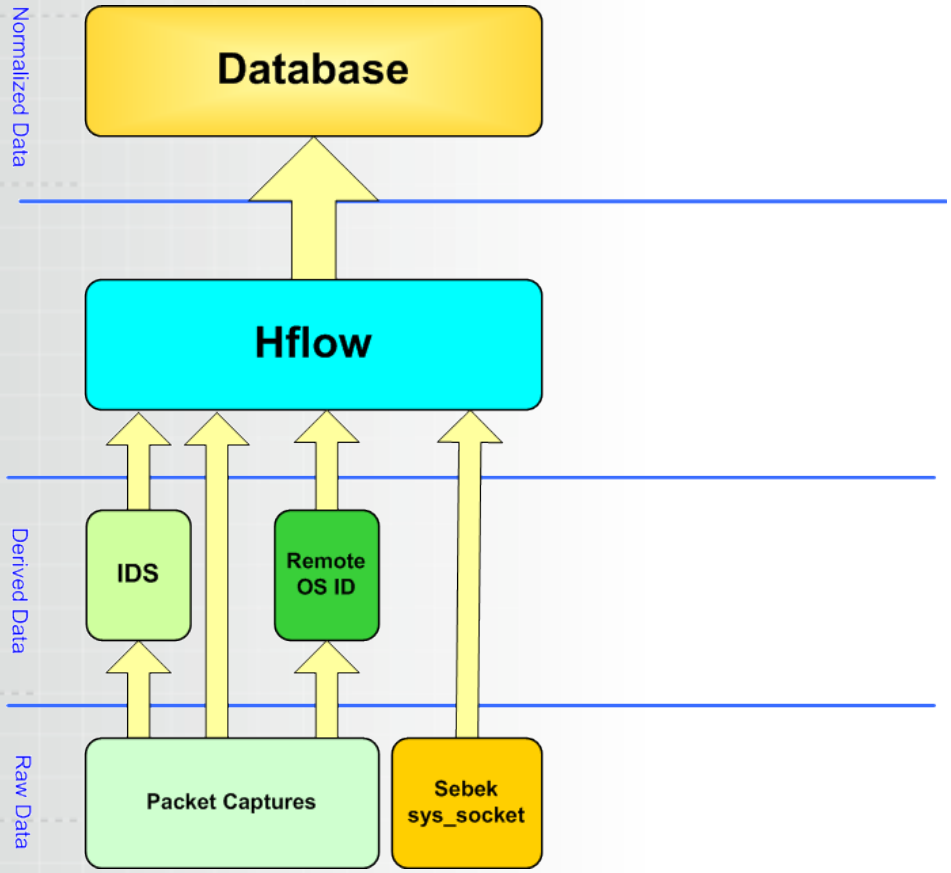


Hfbw Overview

- Fancy perl deamon, which consumes multiple data streams.
- Automates the process of Data Coalescing.
- Inputs:
  - Argus data
  - Snort IDS events.
  - Sebek socket records.
  - p0f OS fingerprints.
- Outputs:
  - normalized honeynet network data uploaded into relational database.

pervasivetechnologylabs
AT INDIANA UNIVERSITY
www.pervasivetechnologylabs.iu.edu

Hflow Illustration



pervasivetechnologylabs
AT INDIANA UNIVERSITY
www.pervasivetechnologylabs.iu.edu

What this gives us.

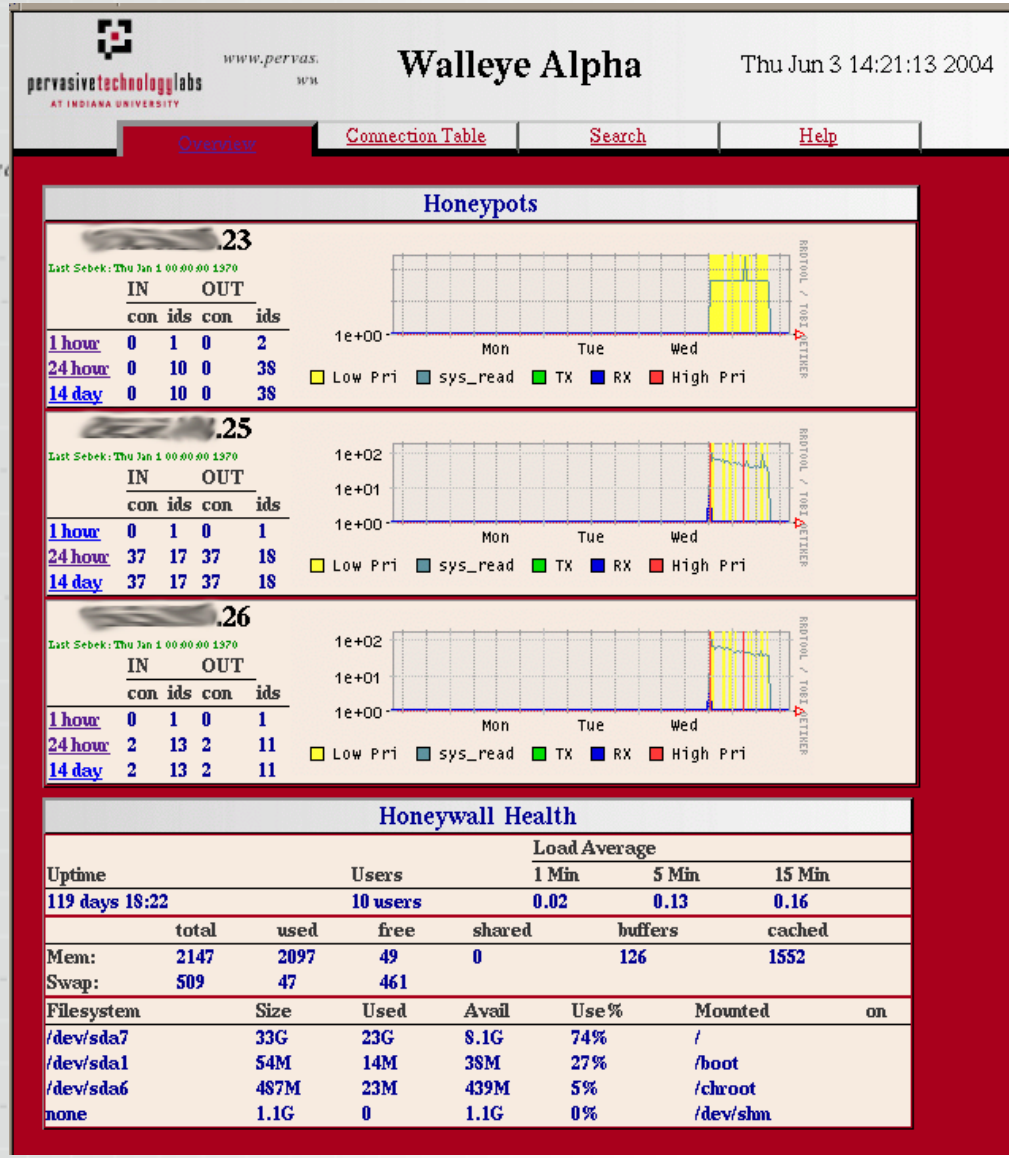
- Automatic identification
  - Type of OS initiating a network connection
  - IDS events related to a network connection
  - IDS evens related to a process and user on a host.
  - Point where non root user gained root access.
  - List of files associated with an intrusion
  - Sense of Attribution between 2 related flows on a monitored box.
- Operate at higher lever where we can scale to support operational networks
  - using Argus central theme of an event sequence can be identified without having to examining packet traces.
  - When packet traces needed, argus info helps facilitate retrieval.

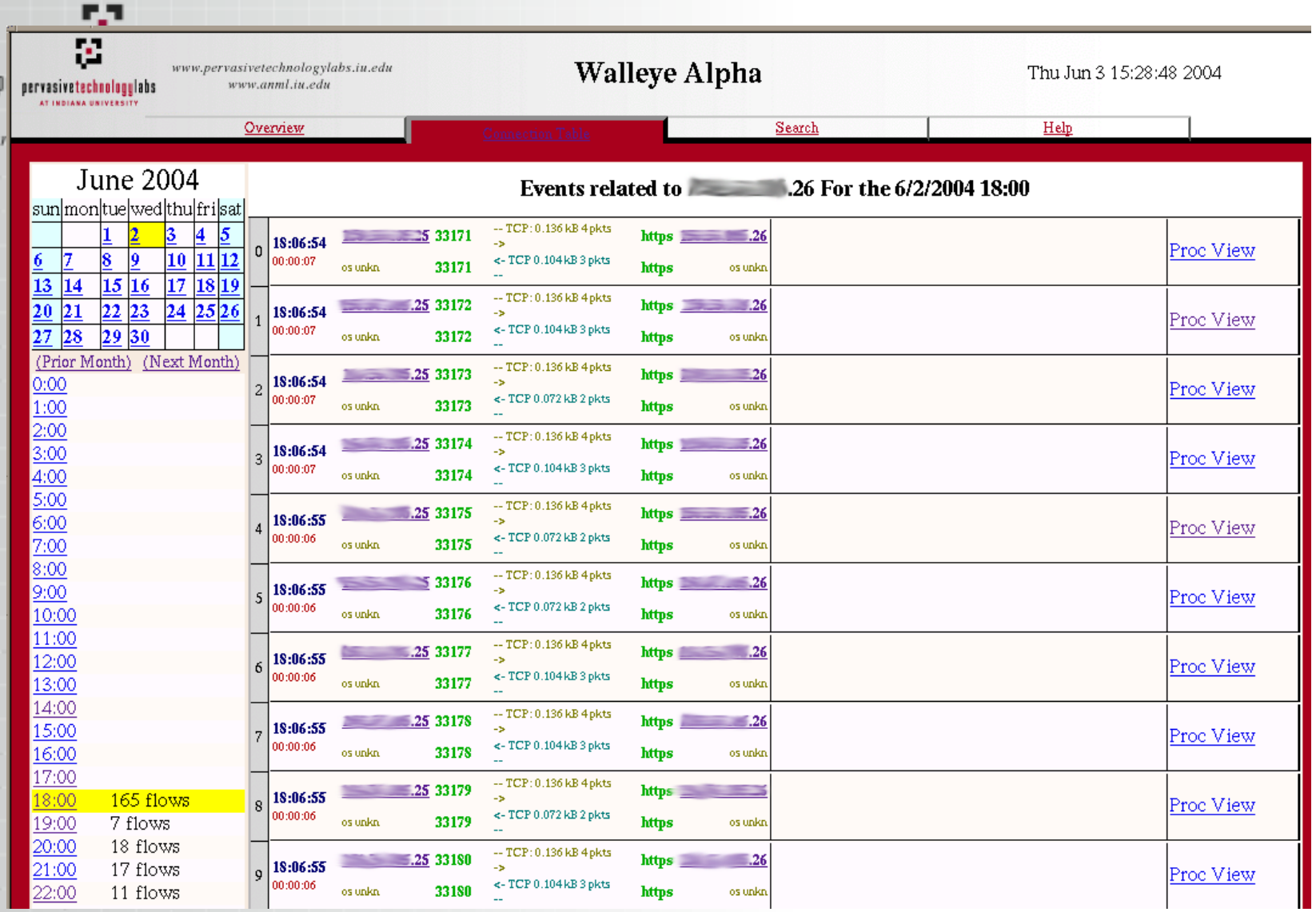
Reporting with Walleye

- perl based web interface
- provides unified view
  - Network “flow” connection records
  - IDS events
  - OS Fingerprints
- Allows user to jump from network to host data.
- Visualizes multiple data types together.

pervasivetechnologylabs
AT INDIANA UNIVERSITY

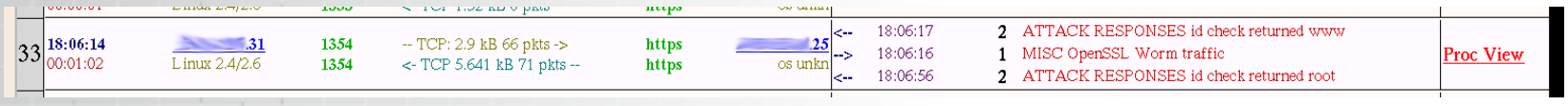
www.pervasivetechnologylabs.iu.ed



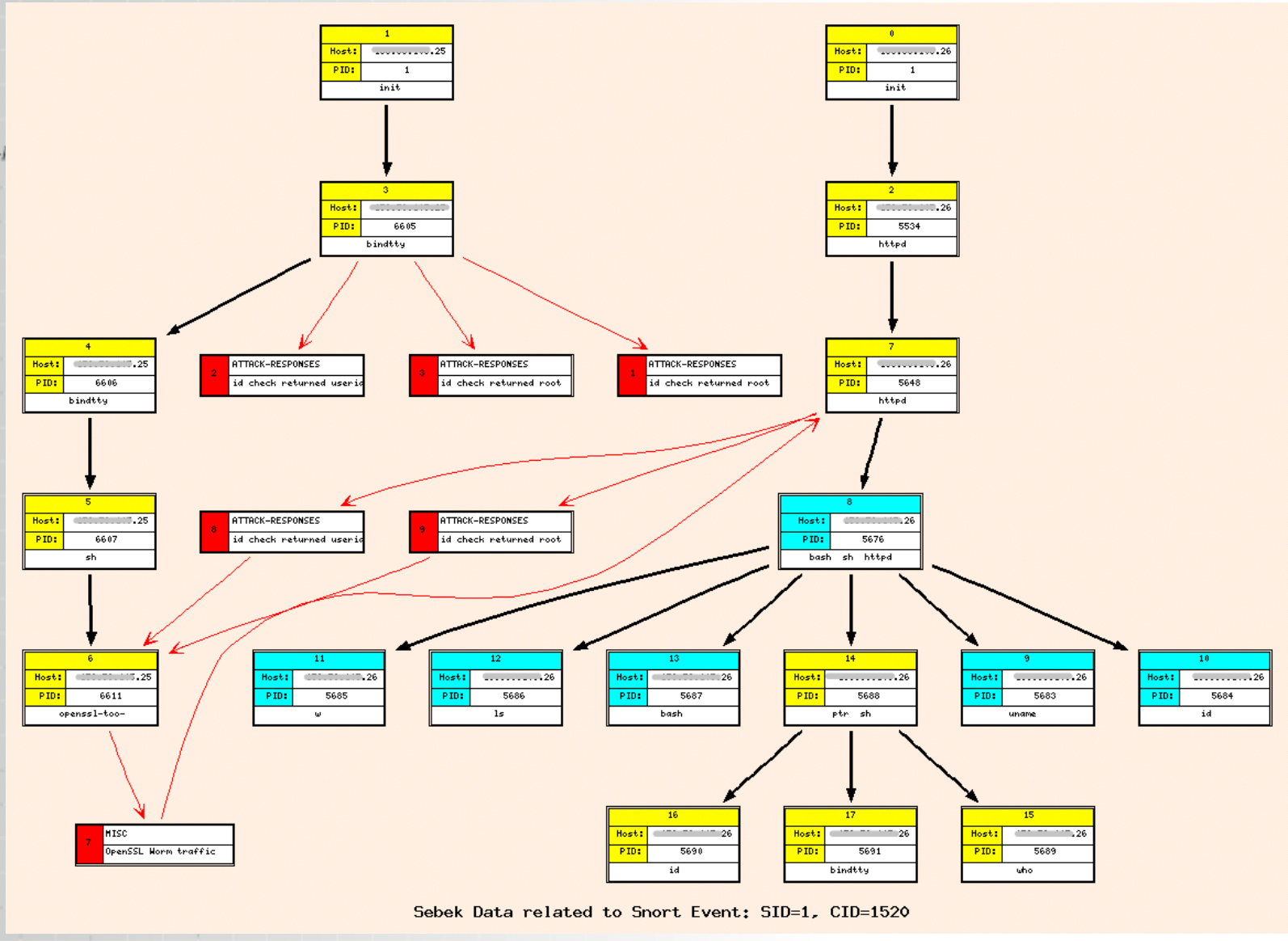

pervasivetechnologylabs
AT INDIANA UNIVERSITY
www.pervasivetechnologylabs.iu.edu
www.anml.iu.edu
Walleye Alpha
Thu Jun 3 15:28:48 2004
Overview
Connection Table
Search
Help
June 2004
sun mon tue wed thu fri sat
1 2 3 4 5
6 7 8 9 10 11 12
13 14 15 16 17 18 19
20 21 22 23 24 25 26
27 28 29 30
(Prior Month) (Next Month)
0:00
1:00
2:00
3:00
4:00
5:00
6:00
7:00
8:00
9:00
10:00
11:00
12:00
13:00
14:00
15:00
16:00
17:00
18:00 165 flows
19:00 7 flows
20:00 18 flows
21:00 17 flows
22:00 11 flows
Events related to .26 For the 6/2/2004 18:00
0 18:06:54 00:00:07 .25 33171 -- TCP: 0.136 kB 4 pkts -> <- TCP 0.104 kB 3 pkts -- https .26 https os unkn os unkn 33171 Proc View
1 18:06:54 00:00:07 .25 33172 -- TCP: 0.136 kB 4 pkts -> <- TCP 0.104 kB 3 pkts -- https .26 https os unkn os unkn 33172 Proc View
2 18:06:54 00:00:07 .25 33173 -- TCP: 0.136 kB 4 pkts -> <- TCP 0.072 kB 2 pkts -- https .26 https os unkn os unkn 33173 Proc View
3 18:06:54 00:00:07 .25 33174 -- TCP: 0.136 kB 4 pkts -> <- TCP 0.104 kB 3 pkts -- https .26 https os unkn os unkn 33174 Proc View
4 18:06:55 00:00:06 .25 33175 -- TCP: 0.136 kB 4 pkts -> <- TCP 0.072 kB 2 pkts -- https .26 https os unkn os unkn 33175 Proc View
5 18:06:55 00:00:06 33176 -- TCP: 0.136 kB 4 pkts -> <- TCP 0.072 kB 2 pkts -- https .26 https os unkn os unkn 33176 Proc View
6 18:06:55 00:00:06 .25 33177 -- TCP: 0.136 kB 4 pkts -> <- TCP 0.104 kB 3 pkts -- https .26 https os unkn os unkn 33177 Proc View
7 18:06:55 00:00:06 .25 33178 -- TCP: 0.136 kB 4 pkts -> <- TCP 0.104 kB 3 pkts -- https .26 https os unkn os unkn 33178 Proc View
8 18:06:55 00:00:06 .25 33179 -- TCP: 0.136 kB 4 pkts -> <- TCP 0.072 kB 2 pkts -- https https os unkn os unkn 33179 Proc View
9 18:06:55 00:00:06 .25 33180 -- TCP: 0.136 kB 4 pkts -> <- TCP 0.104 kB 3 pkts -- https .26 https os unkn os unkn 33180 Proc View

pervasivetechnologylabs
AT INDIANA UNIVERSITY
www.pervasivetechnologylabs.iu.edu

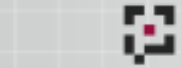
Looking closely



- host x.x.x.31 attacked x.x.x.25 on its https port.
- x.x.x.31 was a linux host.
- The attack matched the OpenSSL worm signature and and triggered 2 additional alerts that indicate the attacker gained www and then root access.
- If we click on Proc View, we jump to a high level view of related process activity.



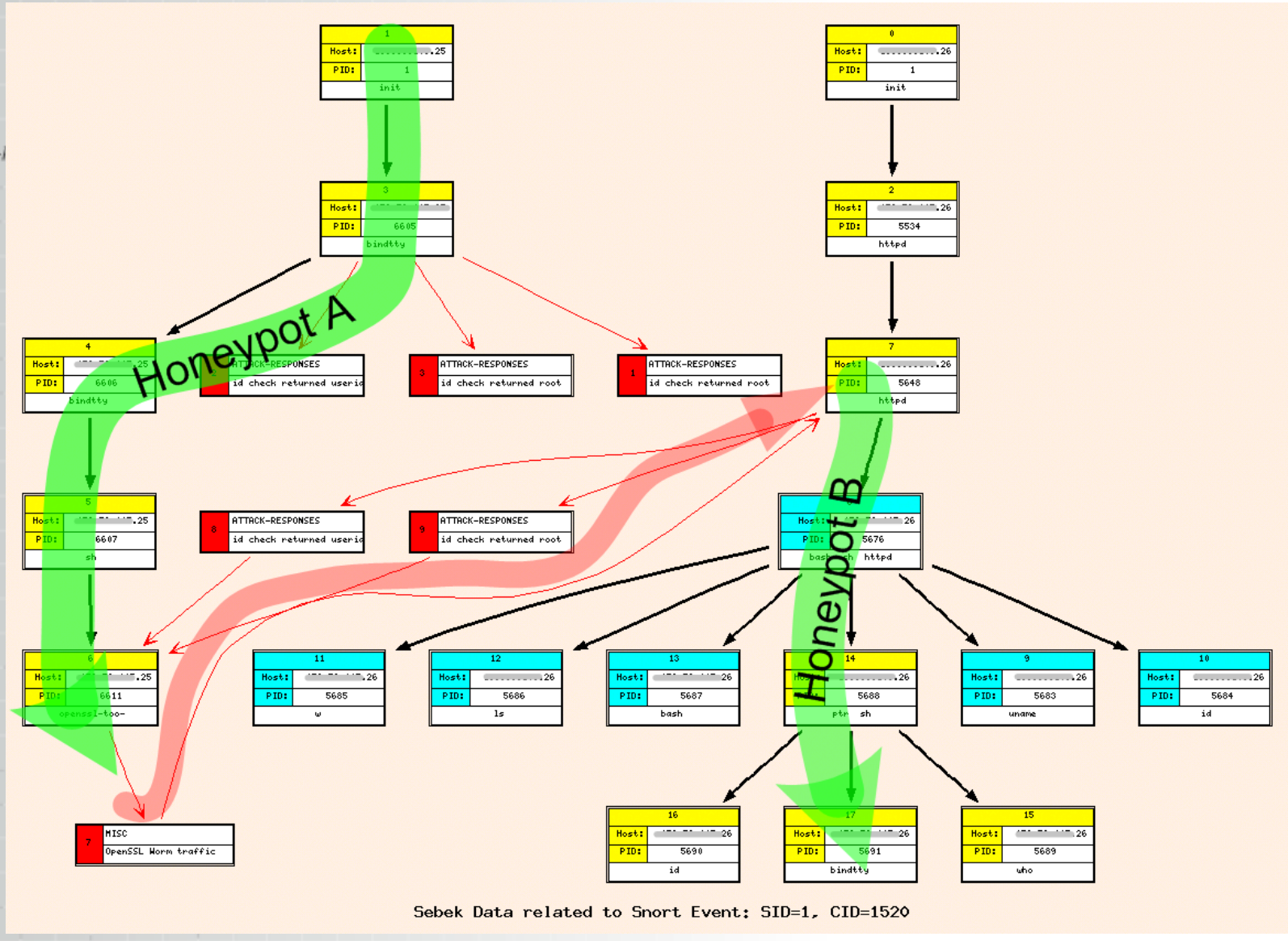
Sebek Data related to Snort Event: SID=1, CID=1520



pervasivetechnologylabs
AT INDIANA UNIVERSITY
www.pervasivetechnologylabs.iu.edu

What you are seeing

- Display shows a process tree and its associated IDS events.
  - created by querying on a single IDS event.
  - Yellow Boxes are root processes
  - Cyan Boxes are non-root processes
  - Red Boxes are IDS events
  - Red Arrow represents direction of flow associated with event
    - Only displaying IDS related flows.
- Graph automatically generated from DB with graphviz tool from ATT.
- Notice anything odd about the graph?



Sebek Data related to Snort Event: SID=1, CID=1520

pervasivetechnologylabs
AT INDIANA UNIVERSITY
www.pervasivetechnologylabs.iu.edu

Walleye tracked intrusion across 2 honeypots

- Both the .25 and .26 honeypots were running the enhanced version of Sebek.
- We are able to provide a sense of attribution in situations where all stepping stones are running Sebek.
- Based on fault tree we could then click on a yellow box and then jump into the sebek interface.

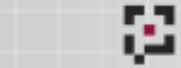
pervasivetechnologylabs
AT INDIANA UNIVERSITY
www.pervasivetechnologylabs.iu.edu

Old question made easy

- What happened after the intrusion?
  - Use IDS event as index into process tree.
  - All related flows will be liked to that tree
  - All files "touched" as part of the intrusion will be related to that tree.
  - Sequences that span 2 hosts can be automatically identified via common network connection.

Features

- Identify descendant f lows or sebek events related to a given event.
- Identify ancestral f lows or sebek events related to a given event
- Effectively, the combination of the two allow us to f lter all data which can not be related to an event of interest.
- Find all f les opened by any process in a process tree.



pervasivetechnologylabs
AT INDIANA UNIVERSITY
www.pervasivetechnologylabs.iu.edu

Current Status

- Sebek
  - socket code in linux client rather stable
  - parent PID tracking currently missing some data for processes that fork and don't read(easy to fix)
- Hf bw
  - few bugs and its not syslog friendly
- Walleye interface
  - a few bugs, look and feel not 100% happy with
  - not yet integrated with conventional analysis tools.
  - doesn't provide way to access raw packets

pervasivetechnologylabs
AT INDIANA UNIVERSITY
www.pervasivetechnologylabs.iu.edu

Future work

- Sebek
  - track fork call so that we always get a view of the process tree
  - look at various anti-anti-sebek options.
- Hf bw
  - testing, lots of testing.
  - evaluate attack resistance
- Walleye
  - get UI to better support workf bw
  - provide alerting
  - provide some summary reports
  - clean, debug, document
  - integrate with existing tools where sensible.
- Get everything to work on the Honeywall CDROM!

pervasivetechnologylabs
AT INDIANA UNIVERSITY
www.pervasivetechnologylabs.iu.edu

Taking this out of the Honeynet context

- Sebek is a good tool for post intrusion intelligence gathering on an intruder
- On a production box it generates great amounts of data, making it diff icult to use.
- With previously mentioned enhancements, Sebek may be a more viable tool, due to its improved coalescing and screening.
- The ability to relate 2 f bws to and from a host via a common process tree may be more valuable than the ability to record keystrokes?

Related works

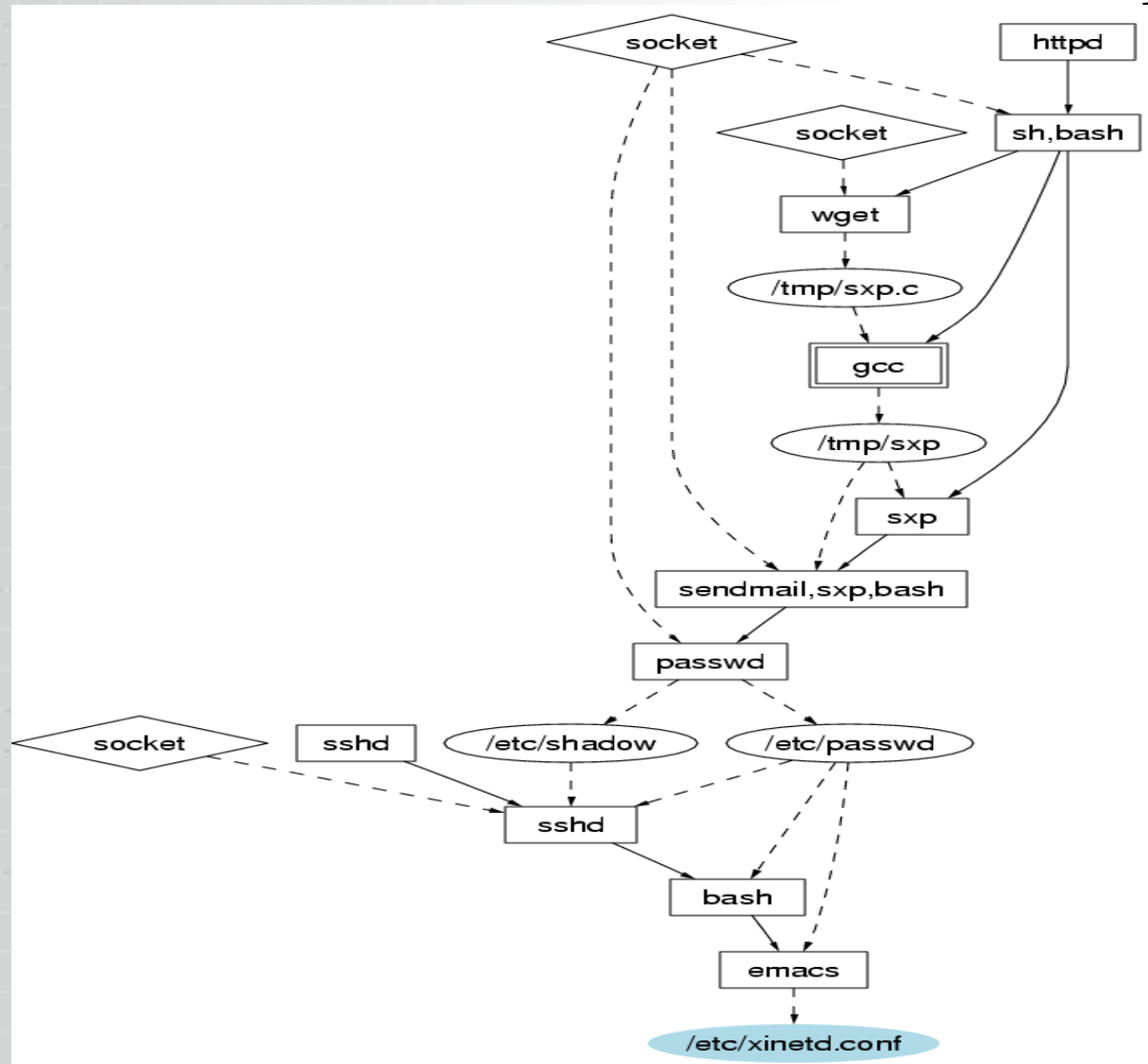
- Covert
- Anti Sebek foo

CoVirt

- CoVirt and the BackTracker system
  - Enhanced UML system allows host to monitor guests system call activity.
  - “Automatically identifies potential sequences of steps that occured in an intrusion.”
  - • Samuel T. King, Peter M. Chen, "Backtracking Intrusions", *Proceedings of the 2003 Symposium on Operating Systems Principles (SOSP)*, October 2003. Award paper.

pervasivetechnologylabs
AT INDIANA UNIVERSITY
www.pervasivetechnologylabs.iu.edu

BackTracker output



References to attack techniques:

- **M. Dornseif, T. Holz, C. Klien, "NoSEBrEak - Attacking Honeypots", Proceedings of the 2004 IEEE Workshop on Information Assurance and Security.**

- **J. Corey, "Advanced Honeypot Identif ication" Jan 2004, http://www.phrack.org/fakes/p62/p62-0x07.txt**