

Filtering Data from Log Files



Andrew Mallett

Author and Trainer

@theurbanpenguin | www.theurbanpenguin.com



Overview



Reading Log Files

- Journalctl
- Root SSH login attempts
- Invalid users
- Listing unique occurrences



awk vs proprietary log filtering tools



**Learning AWK means you
can get the data you need
without paying for
proprietary software**



**For these demos I use a
public AWS Ubuntu system.
It will have been running
just over one day**



```
$ sudo journalctl -u ssh  
$ sudo journalctl -u ssh -g "Failed password for invalid user "  
$ sudo journalctl -u ssh -g "Failed password for root from "
```

Journal Log

On modern Linux systems we have a unified tool to read logs. We can print just entries from the SSH service unit if we want. On this public system there will be many attempts to log in as root or invalid users




```
$ vim ssh-simple.awk
/Failed password for invalid user/{
    print $13
}
/Failed password for root from/{
    print $11
}
$ sudo journalctl -u ssh | awk -f ssh-simple.awk
```

Combine Searches Using AWK

We can combine the searches using awk. The IP address is either field 11 or 13 which we can cater for. We can pipe to sort to gain unique IPs.



Demo



Working with the Journalctl

- Print SSH activity
- Filter with -g
- Filter with AWK




```
$ vim ssh-sort-root.awk
/Failed password for root from /{
    ips[$11]=1
}
END {
    n=asorti(ips, sorted_ips)
    for(i=1; i<=n; i++){
        print sorted_ips[i]
    }
}
```

Sorting Using AWK

The **asorti** function is an internal function in awk that sorts the indices of an array numerically or lexicographically and returns the number of elements in the sorted array.



Demo



Sorting Using AWK

- Sort root logins



Demo



Sorting Using AWK

- Sort root logins and invalid users





Using AWK Output



```
$ sudo apt install -y firewallld  
$ for i in $(sudo journalctl -u ssh | awk -f ssh-final.awk); do  
sudo firewall-cmd --zone=drop --add-source=$i  
done
```

Blocking IPs

If we want to assume these IP addresses should not be on our system we could block each IP with a firewall. We may want to double check that our own IP address is not included, but this shows the use of AWK and filtering



Demo



Blocking Malicious Activity

- Install firewall
- Block IPs listed in log



Summary



AWK and Logs

- Journalctl
 - -u to filter on unit
 - -g to search with grep
- Sort
 - We can simply pipe to sort -u
- All in AWK
 - asorti function can sort array indices
 - The associative index name is the IP address that we can print

Up Next:

Congratulations

