```
*******************************************************************************
***                    UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO                ***
***             DIRECCION GENERAL DE SERVICIOS DE COMPUTO ACADEMICO            ***
***                   DEPARTAMENTO DE SEGURIDAD EN COMPUTO                     ***
***                                UNAM-CERT                                   ***
*******************************************************************************

*******************************************************************************
***                PROYECTO FINAL DEL CURSO PROGRAMACION EN SHELL              ***
***                 ANALIZADOR DE BITACORAS DE CORREO ELECTRONICO               ***
***                       MAIL LOG ANALIZER V 1.0(MLA)                         ***
*******************************************************************************
```

INDICE

1.- INTRODUCCION

El proyecto final del curso de Programación de Shell del programa de Becarios de Seguridad en Cómputo , consiste en una herramienta capaz de analizar y procesar bitácoras de correo electrónico.
La herramienta desarrollada Mail Log Analizer (MLA) en su versión 1.0, permite obtener reportes con información sobre aspectos de seguridad y estadísticos tales como:

1.- Numero de correos Recibidos
2.- Numero de correos escaneados por el antivirus
3.- Numero de correos con algún tipo de Malware
4.- Numero de correos escaneados por el ANTISPAM
5.- Numero de correos detectados como SPAM
6.- Promedio de calificación de correo SPAM
7.- Promedio de calificación de correo NO SPAM
8.- Nombre y cantidad de incidencias de criterios de SPAM

La herramienta tiene un uso sencillo, tanto la instalación como su uso. El archivo de datos que genera cumple con los requisitos del Proyecto, es decir, extrae la información y la separa por campos (los antes mencionados) organizándolos por periodos para cada hora del día.

2.- REQUISITOS DEL SISTEMA

La herramienta MLA ha sido desarrollada en lenguaje de programación de SHELL, tanto Bash como C shell. Por esto los requisitos para la ejecución de la herramienta son:

- Bash o Bourne Shell
- C shell

3.- MODULOS DE LA HERRAMIENTA MLA

La herramienta MLA esta conformada por dos módulos generales cada uno de ellos dependiente de varios archivos con script de shell y de awk.

La estructura general de la Herramienta es la que sigue:

```
-----------
INSTALACION

instalar
|------ mla_calendar
|      | ------ calendar


---------
OPERACION

load
|
|------ inicia
|      |------- logs_ordena.csh
|      |------- mla
|      |       |------- hits_busca.csh
|      |       |       |------- hits_promedio_SI.awk
|      |       |       |------- hits_promedio_NO.awk
|      |       |------- criterios_busca.csh
|      |       |       |------- criterios_ordena.awk
|      |       |       |------- criterios_cuenta.awk
|----- reportes_inicia
|      |------ reportes_mla
|      |       |------- reportes_suma.awk
|      |       |------- reportes_promedio.awk
|      |       |------- reportes_ord_spam.awk
|      |       |------- reportes_cuenta_spam.awk
```

El funcionamiento interno de la herramienta es el que sigue:

instalación
(instalar) : Recibe como entrada los directorios de instalación de la herramienta. Verifica los permisos y efectúa la copia de los archivos necesarios para el funcionamiento de la mismo.

Operación
(load): Modulo maestro el cual recibe los argumentos de entrada de la herramienta. Este es el que ejecuta el modo de operación de la herramienta según los argumentos ingresados (Procesamiento de bitácoras o Reportes).

Básicamente el contenido de este script evalúa los argumentos ingresados, es decir, facilita que el usuario no tenga que ingresar los argumentos en un ORDEN ESPECÍFICO. Con esto la herramienta funcionara correctamente siempre y cuando los argumentos sean valido; en caso de error se mostrara un mensaje de ERROR. Por ejemplo:

```
./load -p              ---    Inicia procesamiento de bitácoras
./load -r              ---    Genera Reporte General
```

Ejemplos y explicación con mayor detalles en Sección 4 (Explicación de Uso)

(inicio): Modulo que ejecuta el procesamiento de la bitácoras de correo electrónico. El procedimiento que se sigue para la generación del archivo de datos es el siguiente:

1.- Lee el todos los archivos del directorio de bitácoras, según la fecha de la primer línea de cada uno, decide cual se procesara primero. Una vez teniendo el orden, separa los sucesos de cada archivo para cada hora del día (0 Hrs-23 Hrs) y crea un archivo temporal para cada hora (horax.tmp) dentro del directorio tmp/ en el directorio de instalación de MLA.

2.- Una vez generados los archivos para cada hora, los verifica cíclicamente para obtener posibles registros para días diferentes, es decir, primero se dividió por hora, después, si existen registros para una misma hora X en un día A y en un día B, divide la hora X para día A y B, generando otros archivos temporales según los días diferentes en esa hora (horaXdiaA.tmp, horaXdiaB.tmp, etc...)

3.- Terminado esto ya se tienen divididos por hora y por día los registros sin procesar, por lo que ahora solo se ejecuta el script mla sobre cada uno de estos archivos. Este script generara un archivo de salida (.out) de cada archivo temporal generado, de los cuales se tomaran cada una de las líneas que se agregaran al archivo final de datos de salida.

(reportes_inicio)

Modulo que genera los reportes personalizados basándose en el archivo de datos generado en la fase de procesamiento de las bitácoras. La finalidad de este modulo es de separar del archivo de datos de procesamiento de bitácoras solo las líneas en donde se tenga coincidencia con los criterios ingresados como argumentos de fechas de inicio, final y nombre de archivo de salida de los reportes. Esto lo hace mediante el numero de mes y aplicando filtros. Una vez que ha generado un archivo con las líneas especificas de fechas , se ejecuta sobre este archivo el script reportes_mla, que da como resultado la generación del archivo del reporte. El funcionamiento interno del script se basa en el uso de awk para el manejo de cada uno de los campos de las líneas del archivo de datos, al que básicamente solo se obtienen sumas y promedios.

(mladirs.conf)

Este es el archivo de configuración que se crea mediante el proceso de la instalación de la herramienta . Este archivo contiene las rutas absolutas de los directorios de trabajo (bitácoras , reportes , datos , etc) y los demás módulos de la herramienta se basan en este archivo para colocar tanto los archivos temporales como los archivos de salida de datos.

Se puede modificar manualmente para cambiar los directorios de trabajo una vez que se ha instalado la herramienta, sin embargo se debe tener cuidado de asignar directorios con los permisos de lectura y escritura adecuados para que el funcionamiento de la herramienta sea el correcto.

El contenido del archivo contiene el siguiente formato:

```
------------------------------
 dir_trabajo=ruta_directorio
------------------------------
```

En cuanto a las líneas de opciones de calendarización:

freq=0 Indica que se configura ejecución cada hora . Cuando freq esta con este valor, los campos de "hora" y "mins" no tienen ningún efecto.

freq=1 Indica que se configura ejecución diaria a la hora indicada en los campos "hora" y "mins". Cuando freq este con este valor, deben ser completados los campos de "hora" y "mins" o de lo contrario podría haber funcionamientos inesperados de la herramienta.

En una instalación default el contenido de mladirs.conf es la siguiente:

```
--------------------------------------------------------------------------------
#DIRECTORIO DE BITACORAS
logs=HOME_DEL_USUARIO/mla/logs
#DIRECTORIO DE REPORTES
reportes=HOME_DEL_USUARIO/mla/reportes
#DIRECTORIO DE DATOS
datos=HOME_DEL_USUARIO/mla/datos
#DIRECTORIO DE INSTALACION
mla=HOME_DEL_USUARIO/mla
#FRECUENCIA DE EJECUCION DE LA HERRAMIENTA
freq=1
hora=12
mins=00
--------------------------------------------------------------------------------
```

4.- EXPLICACION DE USO

INSTALACION:
Para la instalación de la Herramienta se siguen los siguientes pasos:

- Desempaquetar el archivo mla_1.0.tar.gz
  $ tar -zxvf mla_1.0.tar.gz
- Ejecutar el script de instalación (instalar)
  $ ./instalar

  El programa de instalación solicitara 5 entradas de datos:
  - Directorio de ubicación de las bitácoras
    Directorio donde se localizaran los archivos de bitácoras que se van a procesar con la herramienta.
    IMPORTANTE : Es esta primera versión se debe especificar a un directorio dedicado a los archivos de bitácoras, es decir, un directorio el cual solo deberá contener dichos archivos que se van a procesar. La herramienta funcionara correctamente siempre y cuando en este directorio no se encuentren archivos con formatos diferentes a los archivos de bitácoras. Si no se especifica ningún directorio, por default se toma ($/mla/logs).
  - Directorio de Reportes
    Directorio el cual contendrá todos los archivos de salida para los Reportes de las bitácoras procesadas. La herramienta verificara si se tienen los permisos adecuados para lectura y escritura. Si no se especifica ningún directorio, se toma default (~/mla/reportes).
  - Directorio de Datos
    Directorio el cual contendrá el archivo general de datos de salida de las bitácoras procesadas. De este archivo se basa la generación de los reportes. También contendrá un archivo de log que tendrá la hora de inicio y término de cada ejecución de la herramienta. Si no se especifica ningún directorio se toma default (~/mla/datos).
  - Directorio de instalación
    Directorio en el que se copiaran todos los archivos necesarios por la herramienta para su funcionamiento. Si no se especifica ningún directorio se toma default (~/mla)
  - Opción de Calendarizacion de la Herramienta
    Se podrá optar la calendarización de la herramienta en dos modos:
    - ejecución cada hora
    - ejecución diaria a una hora determinada

    El programa de instalación generara un archivo para ejecutarlo como cron.

OPERACION

Una vez instalada la herramienta, se podrá ejecutar en sus dos modos de operación mediante "load" . Las opciones disponibles para load son:

-p : Procesamiento de bitácoras
-r : generación de Reportes
  [-fi mm/dd] : Especifica la fecha de inicio para los Reportes
  [-ff mm/dd] : Especifica la fecha de término para los Reportes
  [-o nombre] : Especifica el nombre del archivo de salida para los Reportes
    Si no se especifica ninguno se toma el default Reporte.dat

El diseño de la herramienta permite ingresar argumentos sin un orden específico. MLA automáticamente tomara los argumentos y verificara si son validos. Esto es importante ya que el usuario no se debe preocupar por ingresar primero la fecha inicial seguida de la fecha final, etc, solo antecederá al argumento la opción. Solo se debe cumplir que el archivo de salida se indique al final.
Ejemplo, para obtener un reporte del 28 de Junio al 1 de Julio se puede hacer de las siguientes formas:

```
$ ./load -r -fi 6/28 -ff 7/1 -o junio28-julio1
$ ./load -r -ff 07/01 -fi 06/28 -o junio28-julio1
$ ./load -r -ff 7/1 -fi 6/28 -o junio28-julio1
```

Otros ejemplos para la generación de Reportes son:

```
$ ./load -r
    Reporte general de todo el archivo de datos
$ ./load -r -o ReporteGeneral
    Reporte general pero se le especifica el nombre ReporteGeneral
$ ./load -r -fi 7/1 -o julio-fin
```

```
        Reporte a partir de Jul 1 hasta la ultima fecha en el archivo
$ ./load -r -ff 11/2 -o ini-nov2
        Reporte a partir de la fecha inicial del archivo hasta Nov 2
$ ./load -r -fi 7/2 -ff 5/3
        Devolvería un error ya que la fecha inicial debe ser anterior
        a la fecha final
```

5.- CONSIDERACIONES DE USO Y LIMITACIONES PARA LA VERSION 1.0

Para esta versión se tienen las siguientes consideraciones importantes y limitaciones para el buen funcionamiento de la herramienta

- En el momento de instalación, asegurarse de que el directorio de bitácoras sea un directorio dedicado a las bitácoras que se procesaran. No importa el numero de bitácoras a procesar, el proceso cíclico de la herramienta se encargara de ordenar y procesar todas,sin embargo, si en el directorio existen otros archivos que posean un formato diferente al de las bitácoras, se podrían tener resultados inesperados en el funcionamiento de la herramienta. Esto se debe a una parte en el código que se basa en el listado del contenido de directorio, y que en una posible versión futura se corregirá.
- En la fase de procesamiento de las bitácoras , en el dado caso de que para una hora determinada no existan registros coincidentes , la herramienta mostrara un mensaje de error de awk , sin embargo NO AFECTA en nada los resultados del archivo de datos procesados , es simplemente una omisión de los mensajes de error dentro del código que en una posible versión futura se corregirá.
- Para ver el log de la ejecución en el caso de la calendarización , puede visualizar el archivo MailDir en el Home del usuario

---
6.- CREDITOS

Desarrollado por:

DGSCA - UNAM
DSC / UNAM-CERT
Plan de Becarios de Seguridad en Cómputo

Javier Ulises Santillán Arenas
bec_jsantillan@correo.seguridad.unam.mx
jusafing@gmail.com

Fecha: 16/07/07

7.- CORRIDA DE EJEMPLO (OPCIONES DEFAULT)

Para visualizar ejemplos de corrida ejemplos, puede revisar los archivos dentro del directorio Ejemplos/ dentro del directorio de instalación de MLA.