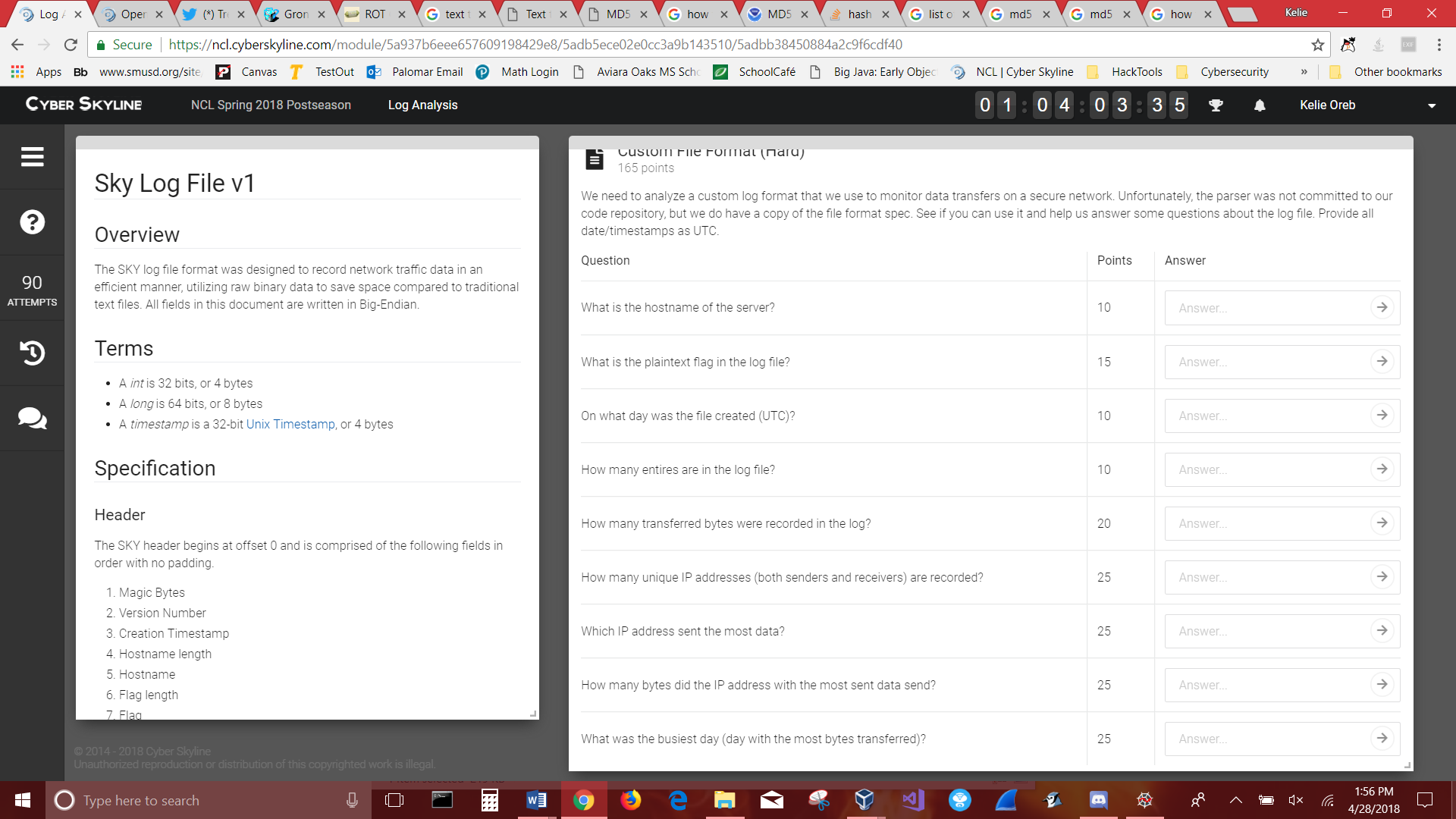
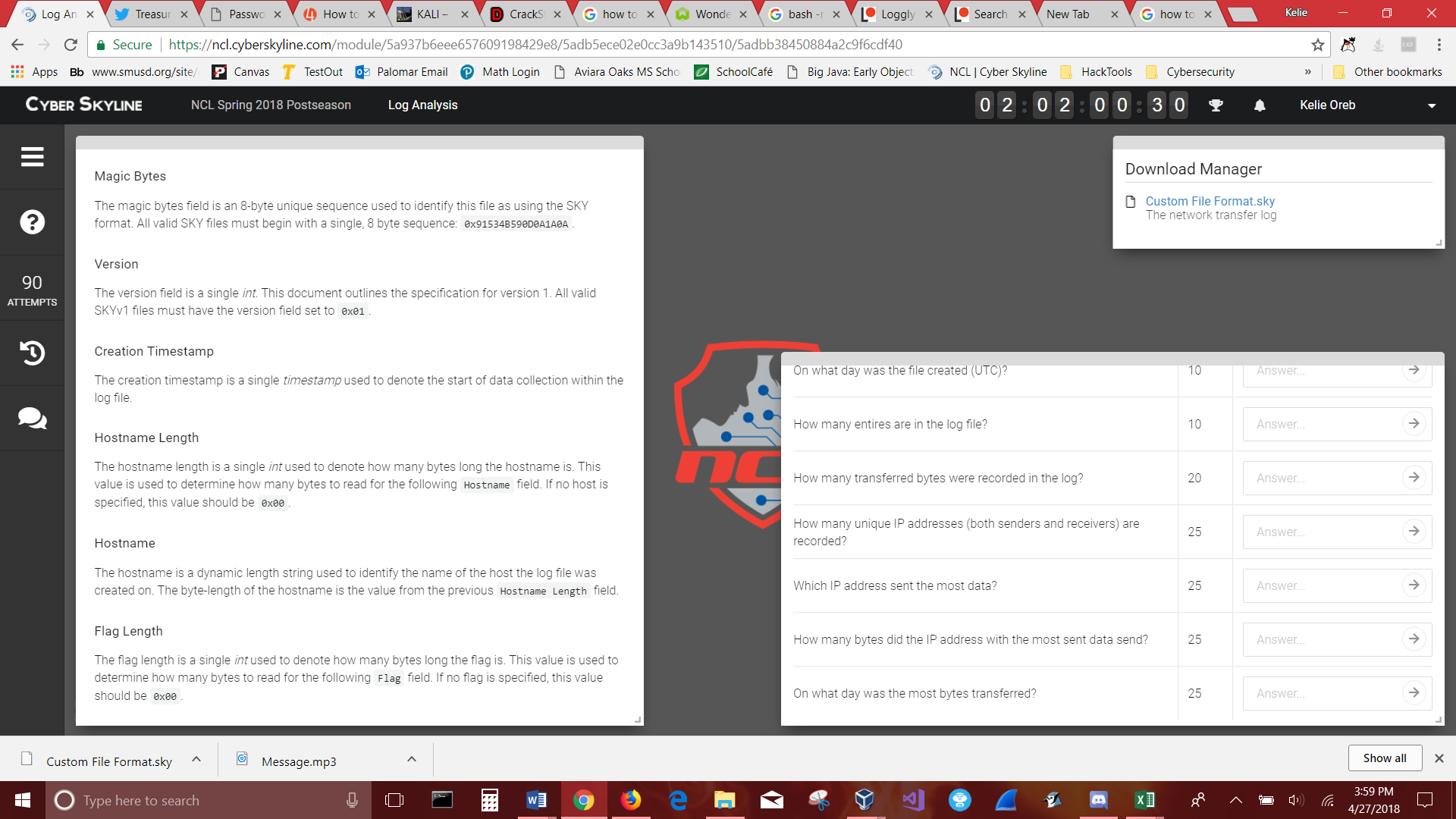
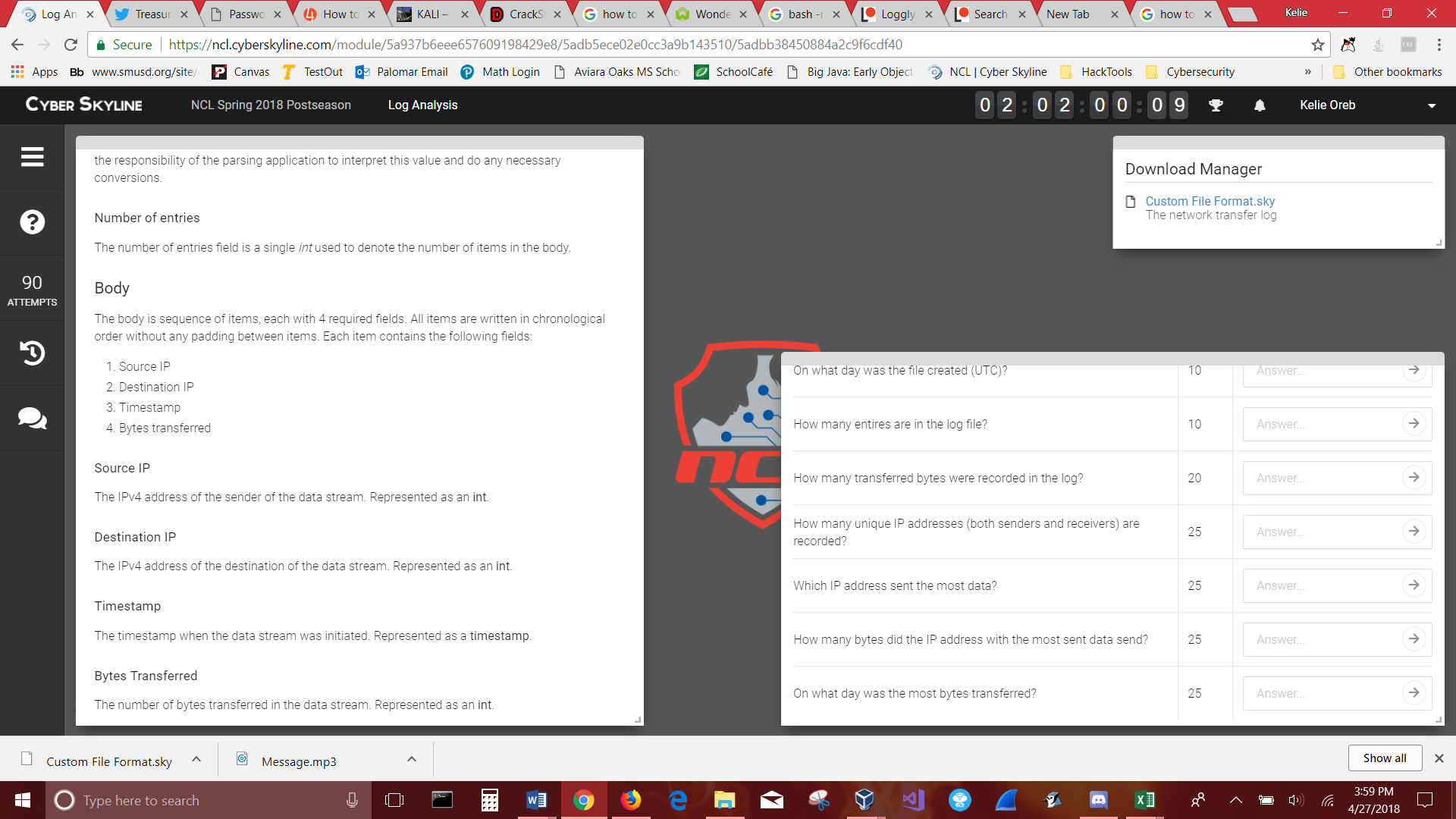
Log Analysis Custom File Format Hard

We need to analyze a custom log format that we use to monitor data transfers on a secure network. Unfortunately, the parser was not committed to our code repository, but we do have a copy of the file format spec. See if you can use it and help us answer some questions about the log file. Provide all date/timestamps as UTC.







Sky Log File v1

Overview

The SKY log file format was designed to record network traffic data in an efficient manner, utilizing raw binary data to save space compared to traditional text files. All fields in this document are written in Big-Endian.

Terms

* A *int* is 32 bits, or 4 bytes
* A *long* is 64 bits, or 8 bytes
* A *timestamp* is a 32-bit [Unix Timestamp](https://en.wikipedia.org/wiki/Unix_time), or 4 bytes

Specification

Header

The SKY header begins at offset 0 and is comprised of the following fields in order with no padding.

1. Magic Bytes
2. Version Number
3. Creation Timestamp
4. Hostname length
5. Hostname
6. Flag length
7. Flag
8. Number of entries

Magic Bytes

The magic bytes field is an 8-byte unique sequence used to identify this file as using the SKY format. All valid SKY files must begin with a single, 8 byte sequence: 0x91534B590D0A1A0A.

Version

The version field is a single *byte*. This document outlines the specification for version 1. All valid SKYv1 files must have the version field set to 0x01.

Creation Timestamp

The creation timestamp is a single *timestamp* used to denote the start of data collection within the log file.

Hostname Length

The hostname length is a single *int* used to denote how many bytes long the hostname is. This value is used to determine how many bytes to read for the following Hostname field. If no host is specified, this value should be 0x00.

Hostname

The hostname is a dynamic length string used to identify the name of the host the log file was created on. The byte-length of the hostname is the value from the previous Hostname Length field.

Flag Length

The flag length is a single *int* used to denote how many bytes long the flag is. This value is used to determine how many bytes to read for the following Flagfield. If no flag is specified, this value should be 0x00.

Flag

The flag is a dynamic length string used to specify a flag value for the log file. The byte-length of the flag is the value from the previous Flag Length field.

**Note**: This field can be used to store both encoded/encrypted flags as well as plaintext flags. It is the responsibility of the parsing application to interpret this value and do any necessary conversions.

Number of entries

The number of entries field is a single *int* used to denote the number of items in the body.

Body

The body is sequence of items, each with 4 required fields. All items are written in chronological order without any padding between items. Each item contains the following fields:

1. Source IP
2. Destination IP
3. Timestamp
4. Bytes transferred

Source IP

The IPv4 address of the sender of the data stream. Represented as an **int**.

Destination IP

The IPv4 address of the destination of the data stream. Represented as an **int**.

Timestamp

The timestamp when the data stream was initiated. Represented as a **timestamp**.

Bytes Transferred

The number of bytes transferred in the data stream. Represented as an **int**.