**Advanced UNIX Scripting Training Hands-on Exercises, Demonstrations & Examples**

**Contents**

[2.0 Python Tooling 3](#_Toc65512622)

[2.1 ArgParser Module 3](#_Toc65512623)

[2.2 File Operations 3](#_Toc65512624)

[2.3 Logging with Python - Basics 3](#_Toc65512625)

[2.4 Python Objects 3](#_Toc65512626)

[2.5 Data Analysis Tools and Techniques 3](#_Toc65512627)

[2.6 Long Tail Analysis: A New Hope in the Cybercrime Battle 4](#_Toc65512628)

[2.7 Viewing Windows Event Logs with Python 5](#_Toc65512629)

[2.8 Packet Analysis 6](#_Toc65512630)

# Python Tooling

## ArgParser Module

Refer to the following book pages on the Mastering Python book: **Pg. 25 - 27**

## File Operations

Refer to the following book pages on the Mastering Python book: **Pg. 52 - 53**

## Logging with Python - Basics

Refer to the following book pages on the Mastering Python book: **Pg. 406 – 413**

## Python Objects

Refer to the following book pages on the Mastering Python book: **Pg. 16 – 17**

## Data Analysis Tools and Techniques

Refer to the following guides and cheat sheets for Data Analysis:

<https://pandas.pydata.org/pandas-docs/stable/user_guide/10min.html>

<https://pandas.pydata.org/Pandas_Cheat_Sheet.pdf>

## Long Tail Analysis: A New Hope in the Cybercrime Battle

Read the following short article after covering the section slides:

<https://threatpost.com/long-tail-analysis-hope-cybercrime-battle/155992/>

The important takeaway points:

A security analyst uses this basic four-step process for long-tail analysis:

1. The analyst finds events of interest, such as website connections or user authentication. Then, you determine how to aggregate the events in a way that provides enough meaning for analysis. As an example, you can graph user accounts by the number of authentication events or web domains by the number of connections.
2. This grouping of data will create a distribution that might be skewed in a particular direction with a long tail, either to the left or right. You might be particularly interested in the objects that fall within that long tail. These are the objects that are extracted, in table format, for further analysis.
3. You then investigate each object as necessary. In the case of authentications, you would look at the account owner, the number of authentication events and the purpose of the account, all with the intended goal of understanding why that specific behavior is occurring.
4. Determine what actions, if any, you need to take and proceed to the next object. You might decide to simply ignore the event and repeat step 3 with the next object. Otherwise, the next steps include working with incident responders or your IT team.

## Viewing Windows Event Logs with Python

Use the following article as a reference for how to open windows event logs with Python. Can you think of ways to apply regex to efficiently sift through logs?

**Challenge: Try to find the critical event. What is the event ID? When did the event occur?**

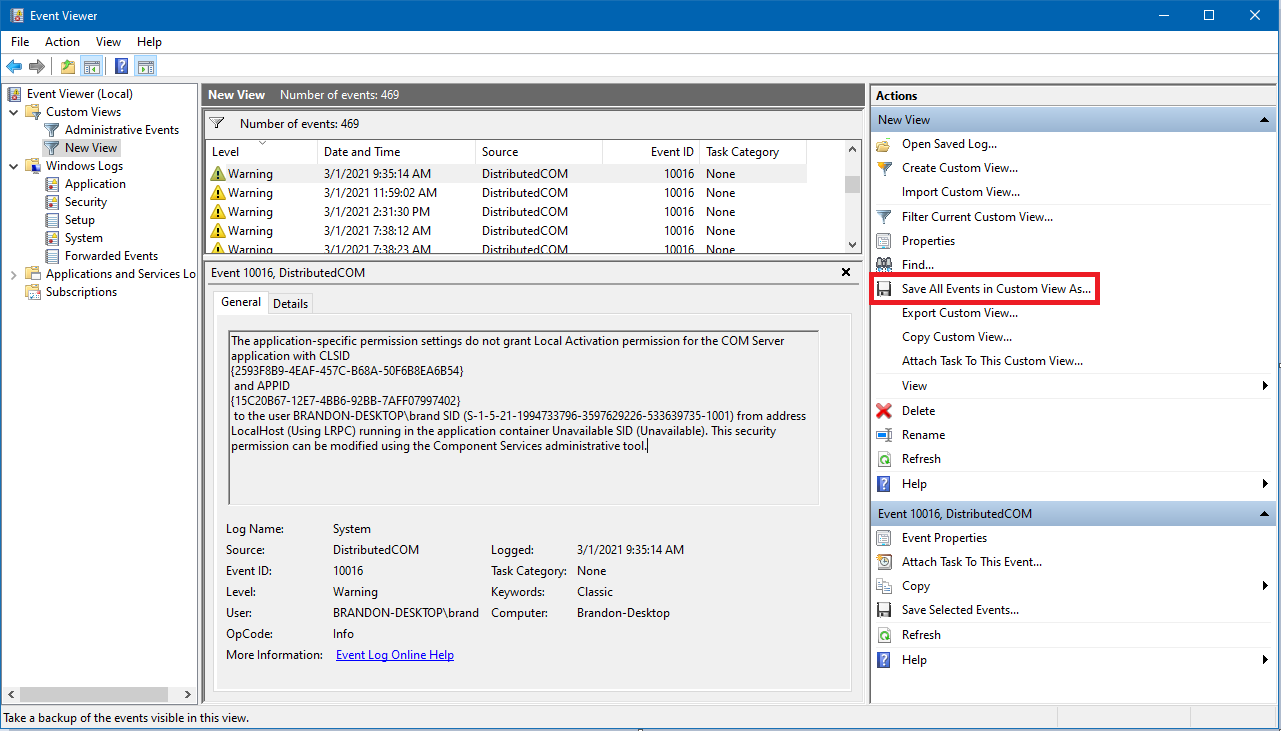
**How many Error logs are present among the 469 events?**

<https://chapinb.com/python-forensics-handbook/ch03_event_logs.html>

Refer to WindowsEvents.evtx provided.

<https://github.com/williballenthin/python-evtx>

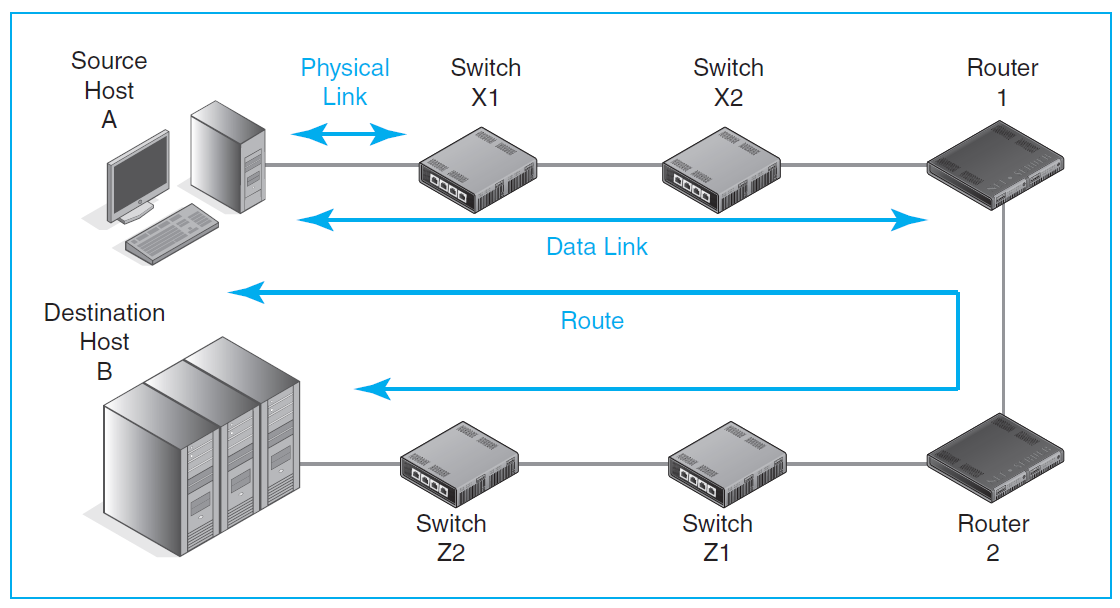
Information about WindowsEvents.evtx:



## Packet Analysis

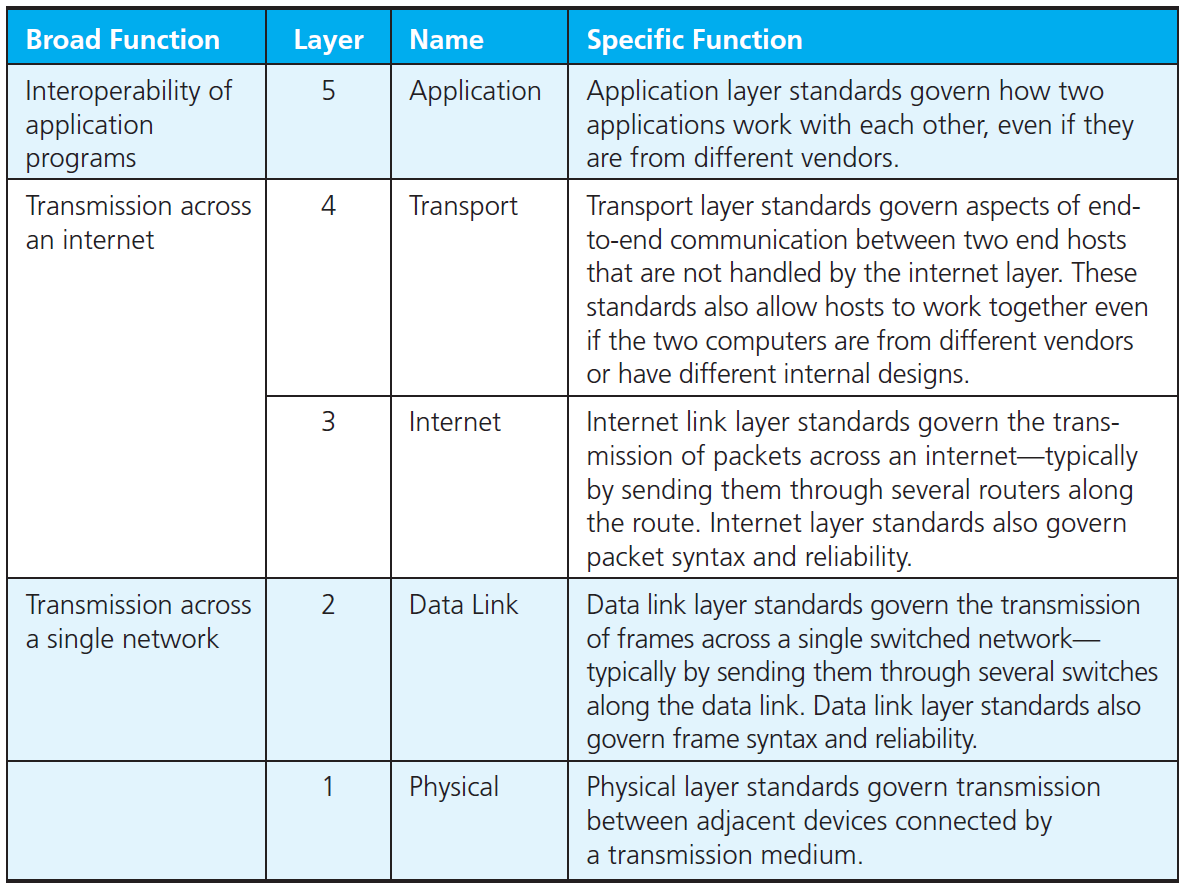
* Physical links are connections between adjacent devices, such as a host and a switch, two switches, two routers, a host and a switch, and so forth.
* Physical layer standards are not concerned with messages. Their job is to turn the bits of data link layer messages (frames) into signals.
* Data link layer standards govern the transmission of frames between two hosts, two routers, or a host and a router across a single point-to-point, switched, wireless, or hybrid switched/wireless network. The path that a frame takes is called its data link. This layer governs switch operation and frame organization.
* Internet layer standards govern the transmission of packets from the source host to the destination host, across multiple networks in an internet. The path that a packet takes between the two hosts is called its route. This layer governs router operation and packet organization.

A common source of confusion is that concepts are repeated at the data link and internet layers but with different terminology. This occurs because internetworking required the adding of a second layer of standards to those needed for transmission through single networks.



* The physical, data link, and internet layers are for standards that move packets along their way between the source host and the destination host. In contrast, Figure 2-4 shows that transport and application processes govern processes that exist only on the two communicating hosts.
* The transport layer supplements the internet layer. Internet layer operation typically is a best-effort service that does not guarantee that packets will be delivered. The transport layer is a “fix-up” layer that can add reliability and other desirable characteristics to transmission across an internet. In addition, the source host transport layer process fragments application messages. These fragments are sent in individual packets. The destination host transport process reassembles the segments and passes the application message to the application. The application layer is for application standards. When two e-mail programs need to work together, they use an e-mail application standard. For webservice, HTTP is an application layer standard. There are more application layer standards than there are standards at all other layers combined because there are so many applications and because different applications usually need different application standards.

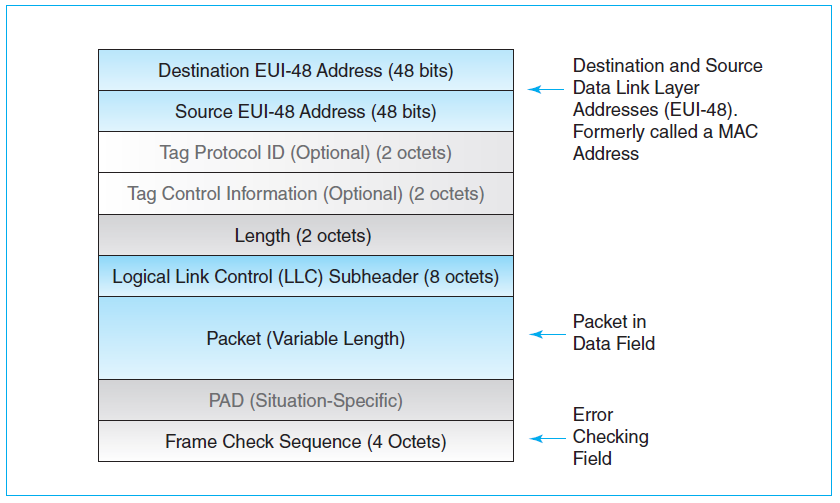
The Five Layers Figure below recaps the five layers.



Speaking broadly, the physical and data link layers govern transmission through single networks. Also speaking broadly, the internet and transport layers together govern transmission

through an internet. The internet layer governs packet organization and raw packet delivery. The transport layer fixes up problems and does fragmentation and assembly. Finally, the application layer governs how two applications work together.

**Ethernet frame**



Messages at the data link layer are frames. In wired local area networks (LANs), the

dominant network standard is Ethernet. Actually, Ethernet, like most “standards,” is

really a family of standards. Ethernet has many different physical layer protocols from

which a company can choose. However, generally speaking, it has a single data link

layer frame standard, which Figure 2-10 illustrates.

The fields in the frame are delimited by their lengths in octets or bits. The destination

host or switch first receives the first bit of the destination address field. It then

counts bits until, 48 bits later, it gets to the next field, then to the field after it, and so

forth. Then it can process the frame.

Ethernet has a complex frame syntax. We will look at its components in more

detail in Chapter 5. There are only four fields that we need to emphasize at this time:

the source and destination address fields, the data field packet, and the Frame Check

Sequence field.

Source and Destination Address Fields Ethernet has a data link layer destination

and address fields. Each address is 48 bits long. Traditionally, these have been

called MAC addresses. However, they are now called Extended Unique Identifier

48-bit (EUI-48) addresses.

**Questions:**

1) Which is the dominant network standard in wired LANs?

2) Name the data link layer address used in this standard. What is the size of this address in bits

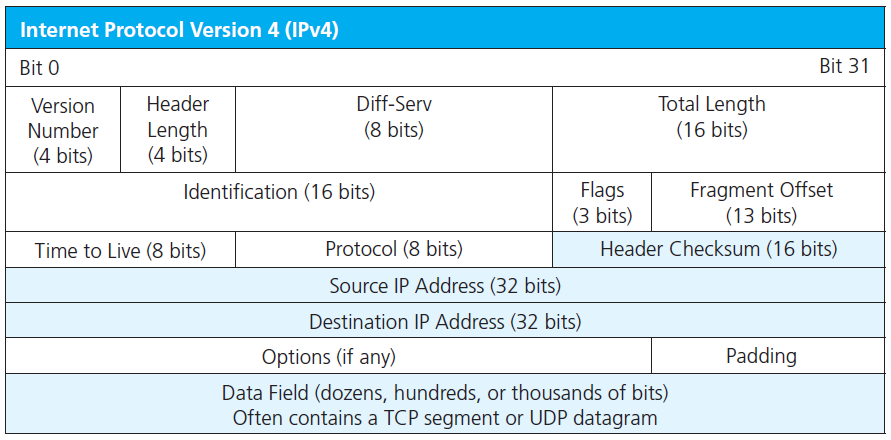
and octets?

3) How do destination hosts identify different fields in the frame of this standard?

4) What is the role of frame check sequence field in error detection?

5) Can Ethernet correct an erroneous frame?

**IP Packet**



An IP packet, like an Ethernet frame, is a long string of bits (1s and 0s). Of course, drawing the packet this way would require a page several meters wide. Instead, we usually depict an IP packet as a series of rows with 32 bits per row. In binary counting, the first bit is zero. Consequently, the first row shows bits 0 through 31. The next row shows bits 32 through 63. This is a different way of showing syntax than we saw with the Ethernet frame, but it is a common way of showing syntax in TCP/IP standards, so you need to be familiar with it.

Each IPv4 packet has source and destination IP addresses. Each is 32 bits long, so each has its own row in the header. Routers use destination IP addresses to decide how to forward packets so that they will get closer to their destination.

The IPv4 **Header Checksum** field is like the Frame Check Sequence field in the Ethernet frame. It also is used for error detection. As in the case of Ethernet frames, incorrect IP packets are simply discarded. There is no retransmission. So like Ethernet, IP is not a reliable protocol.

**Questions:**

1) How many octets long is an IPv4 header if there are no options?

2) What is the bit number of the first bit in the destination address field in IPv4? (Remember that the first bit in binary counting is Bit 0.)

3) How long are IPv4 addresses?

4) What device in an internet besides the destination host reads the destination IP address?

5) What is this device’s purpose in doing so?

6) Is IP reliable or unreliable?