

Python tool to decode base64 encoded data

¹Mukesh Choudhary M. Tech. Student CS department ² Mr. Mohit Jain Asst. Professor CS department

BM College of technology Indore, India. BM College of technology Indore, India.

mukesh.choudhary311@gmail.com* bmctmohitcs@gmail.com**

Abstract:

Base64 encoding is a process of converting binary data to an ASCII string format by converting that binary data into a 6-bit character representation. The Base64 method of encoding is used when binary data, such as images or video, needs to be transmitted over systems that are designed only to transmit data in a plain text (ASCII) format. Web developers use base64 formats to include images, fonts, sounds and other resources directly inside HTML, JavaScript, JSON and XML files. We estimate that billions of base64 messages are decoded every day. Due to wide use of base64 encoding a decoder software is required. In this project, a tool is developed to decode the base64 encoded data. Tool is developed using Python language, which is the most widely used software language these days. Python is used by hundreds of thousands of programmers and is used in many places. Sometimes only Python code is used for a program, but most of the time it is used to do simple jobs while another programming language is used to do tasks that are more complicated. Its standard library is made up of many functions that come with Python when it is installed. Many libraries available make it possible for the Python language to do more things. These libraries make it a powerful language. it can do many different things. For taking the user input, a GUI is developed in python, which takes the input from user and saves to a text file. The python script uses the text file to extract the data and decode the encoded data. As now days many IOT systems are coming up which need to send small packets of data to the server and later server processes it to make a user understandable data. Like the IOT device sends the data of the home appliances in JSON format and sends it after doing base 64 encoding. This tool will help a user to see the actual data. If the data set is of large size ZIP compression/ decompression can be used. In this way working on this project provided a very useful platform learn Python, GUI, base64 encoding decoding etc.

Keywords: XML, JSON, HTML, IOT, ZIP, GUI, ASCII.

1. INTRODUCTION

Base64 theory

Base64 uses 64 characters that are available (and printable) in most text encoding schemes to represent its data. So how to represent a data byte which can hold 256 different values by a set of only 64 characters? Well, of course this is not possible by a 1:1 mapping, so the Base64 encoded data stream needs more characters than the original byte-based. In fact three data byte values are encoded to four Base64 values: This makes sense as three 8bit values are in total 24bit, while a value in range 0-63 requires 6bit for representation, thus four Base64 characters fit in 24bits.

Base64 is an algorithm to convert a stream of bytes into a stream of printable characters (and back). The origin of such binary-to-text encoding scheme like Base64 is the requirement to send a stream of bytes over a communication channel which does not allow binary data but only text-based data. E.g. the ASCII standard is based on seven-bit values, thus uses a range of 0-127. Still an ASCII value is normally stored in a byte which has eight bits (0-255). The usage of the most significant bit of an ascii value is not completely defined (could be used as a parity bit for example). So if a stream which is expected to contain only ASCII values actually also contain non-ASCII values, this could potentially lead to problems. This is where Base64 encoding & decoding could come into play.

JSON

JavaScript Object Notion, which is known as JSON, is one of the most popular data transition formats. It is a text-based and lightweight format for data transactions. Douglas first computed JSON format. This being a text-based format is easier to read or write by the user and at the same time, its lightweight property makes it a stress-free alternative for machines to deconstruct or generate. It is basically a subset of the JavaScript but JSON, as a text format is totally independent of any of the programming languages used as almost all the languages, can easily analyze the text. Its unique properties like text-based, lightweight, language independence etc. make it an ideal candidate for the data-interchange operations.

Why JSON?

Although standardization through Ecma International and the IETF has helped JSON gain industry acceptance, other factors have popularized JSON. The increasing popularity of JavaScript JavaScript's resurgence is boosting JSON's popularity. Over the past several years, we have seen the rise of JavaScript as a first-class development language and environment. This ecosystem includes platforms such as Node.js, and Model/View/Controller (MVC) frameworks such as AngularJS, React, Backbone, and Ember.

The key and their values are grouped using curly braces, both opening and closing "{ }". So, in the previous *case* when we were creating a JSON with a car attribute, we were actually creating a JSON car Object. There are certain rules that need to be followed while creating a JSON structure, we will learn about those rules while discussing the Key value pairs.

So, in order to create a JSON, the first thing we will need is an attribute. Here, we are creating an "Employee" JSON object. Next thing we need is to specify the properties of the object, let's assume our employee have a "First Name", "Last Name", "employee ID" and "designation". These properties of the employee are represented as "Keys" in the JSON structure.

Let us create a JSON object:

```
{
  "FirstName"      : "Sam",
  "LastName"       : "Jackson",
  "employeeID"     : 5698523,
  "Designation"    : "Manager",
}
```

PYTHON

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently whereas the other languages use punctuations. It has fewer syntactical constructions than other languages.

- Python is interpreted: the interpreter processes Python at runtime. You do not need to compile your program before executing it. This is similar to PERL and PHP.
- Python is Interactive: You can actually sit at a Python prompt and interact with the interpreter directly to write your programs.

Python 3 – GUI Programming (Tkinter)

Python provides various options for developing graphical user interfaces (GUIs). The most important features are listed below.

- Tkinter: Tkinter is the Python interface to the Tk GUI toolkit shipped with Python.
- wxPython: This is an open-source Python interface for wxWidgets GUI toolkit. You can find a complete tutorial on WxPython

PyQt: This is also a Python interface for a popular cross-platform Qt GUI library. TutorialsPoint has a very good tutorial on PyQt

- JPython: JPython is a Python port for Java, which gives Python scripts seamless access to the Java class libraries on the local machine

There are many other interfaces available, which you can find them on the net

Tkinter Programming

Tkinter is the standard GUI library for Python. Python when combined with Tkinter provides a fast and easy way to create GUI applications. Tkinter provides a powerful object-oriented interface to the Tk GUI toolkit. Creating a GUI application using Tkinter is an easy task. All you need to do is perform the following steps –

- Import the Tkinter module.
- Create the GUI application main window.
- Add one or more of the above-mentioned widgets to the GUI application.
- Enter the main event loop to take action against each event triggered by the user.

2. RELATED WORK

Various approaches are followed to encode decode base 64 data using different tools.

JavaScript Object Notation (JSON) is a lightweight data-interchange format. It is easy for humans to read and write. It has a data format that is interchangeable with a programming languages built-in data structures that eliminates translation time and reduces complexity and processing time.

Somchai, Wen Dong proposed In order to realize the data encryption of enterprise information platform, the principle of Encryption and Decryption of Base64 algorithm is expatiated and explained in detail. [1]

The particular set of 64 characters chosen to represent the 64 place-values for the base varies between implementations. The general strategy is to choose 64 characters that are both members of a subset common to most encodings, and printable [2].

The new AES will certainly become the de facto standard for encrypting all forms of electronic information, replacing DES. AES-encrypted data is unbreakable in the sense that known cryptanalysis attack can decrypt the AES cipher text without using a brute-force search through all possible 256 bit keys. Security is no longer an afterthought in anyone's software design and development process. It will greatly increase the reliability and safety of your software systems.[3]

Data interchange format have significant consequences on data transferring rates and performance, data interchange format generate from mark-up to further support for structural attribute of information using encoding of meta-data. XML and JSON are data interchange format that can be use in different aspects with unique purpose. XML primary uses are object serialization for transfer of data between application and Remote Procedure Calls RPC [4].

To analyze the impact of management of energy and cost of processing for transferring data in mobile devices of proposed formats of XML, JSON and Protocol Buffer [5]. These two form of data interchangeable using data serialization approach which allows for better

communication between applications. Data transmission of web application more secure, powerful in the XML serialization approach and with JSON serialization approach fast and convenient [6].

XML is more complex than JSON in web services of web programming, some time programmer doesn't need to use namespace and mixed content documents. The developers focus on to use simple data structure, compact and exchange format. XML is great in problem domain, namespace, wellformed and mixed content document [7].

The Analytic Hierarchy Process is a method of measurement for formulating and analyzing decisions. AHP is a decision support tool which can be used to solve complex decision problems considering tangible and intangible aspects. Therefore, it supports decision makers to make decisions involving their experience, knowledge and intuition [8].

Base64 is one of the most common encoding styles to transfer 8 byte code in the internet. Base64 is used widely, through which the data is usually put in URL. However, the mark % as a wildcard will be generated after encoding by standard Base64[9].

Security and confidentiality is one important aspect of an information system [10]. The information sent is expected to be well received only by those who have the right. Information will be useless if at the time of transmission intercepted or hijacked by an unauthorized person [11]. The public network is one that is prone to be intercepted or hijacked [12]. From time to time the data transmission technology has developed so rapidly. Base64 is not an encryption method, but it is the standard encoding [13]. Base64 history begins with electronic mail. At that time, the email is sent by SMTP (simple mail transfer protocol) to a mail server, then sent to the mailbox at the mail server destination. A Protocol is an ordinance of the computer to communicate with each other via a network [14].

Base64 is used commonly in multiple applications including email through MIME and storage of complex data in XML. Base64 needs to be learned because the transformation base64 widely used on the Internet as a medium to transmit data format data. Due to the result of the transformation, base64 be plain text, and then this value will be much more easily shipped, compared to the form of binary data format [15].

In reference to the above work the system is focused to increase its capability and make it a complete data logger unit for electricity power consumptions.

3. Methodology

The system will be developed in view of developing software and testing.

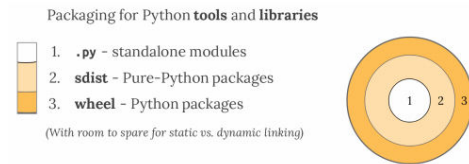
3.1 Implementation plan

1. Understanding the system requirement.
2. Understanding the functionality expected and planning the steps to make it possible.
3. Choosing the programming language finding the best language to complete the project.
4. Understanding the pros and cons of chosen platform.
5. Understanding the V model of software development.
6. Software implementation and testing it with raw data.
7. Testing the UI developed and testing it with different scenario.
8. Integration testing of the system and solving issue found

Package creation and installation

These are just a few of the tools Python's ecosystem provides for distributing Python code to developers, which you can read about in Packaging and distributing projects. The following approaches to packaging are meant for libraries and tools used by technical audience in a development setting. If you're looking for ways to package Python for a non-technical audience and/or a production setting, skip ahead to Packaging Python applications.

Python modules A Python file, provided it only relies on the standard library, can be redistributed and reused. You will also need to ensure it's written for the right version of Python, and only relies on the standard library. Python's sdist's are compressed archives (.tar.gz files) containing one or more packages or modules.



The most basic directory structure for a Python package looks like this

```

project
|
|_ setup.py
|
|_ myPackage
|
|_ somePython.py
|_ __init__.py
    
```

4. PRACTICAL MODEL

A work flow of a system is the stepwise representation of the operation of the system.

1. Start
2. Initiate the system and its different modules.
3. Initiate the UI input function to take input continuously.
4. Initiate the base64 module and import the library.
5. Write the program for taking input from user.
6. Write the program to process the input.
7. Write the program for printing the result on UI.
8. Make the program a python module.
9. Make it a ZIP installable folder.
10. Install the module.
11. Execute the program and put the user input.
12. Take the result and analyze the result.
13. Fix if anything is wrong and re test.

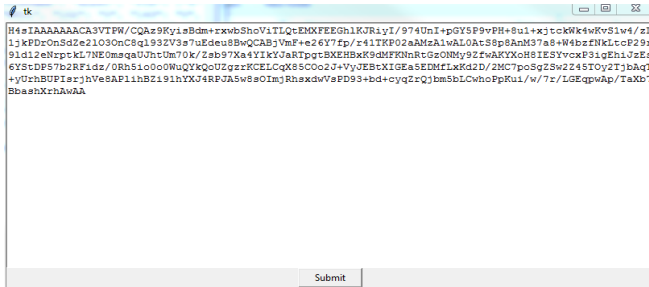
4.4 Results

Step 1. Here we have a base64 encoded data (A real time data)

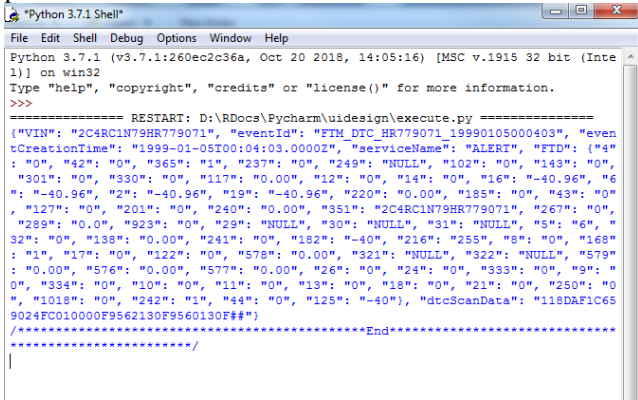
H4sIAAAAAACA3VTu27bMBT9IU Ae2xj3yUc2V7bQ AKmH1O3QJTASDRmaAqmRJci/95KUJbJBjvEcnvs6l3r tfl7vu6uLjnjq57XHv49db7yN47D5fdOPL+HS6fkj3w+Hb 3fbQ352v7wgwgiKASGCe1f3zeDw9/nk6PP4ec16TXYJe IhxAriRcMa8BRX6liL/j88vj/bg/Fu3mZnd7SPxw2Bp+7S SxuROhdNz/uLIJiJ0mCOIM7OczLqzEOgCBFpHwnJZhC WGGRYM5qYWtoR

Step 2. Open the tool by opening the execute.py file in python.

Step 3. Go to Run option and Run Module below UI will prompt out. Paste the raw data in the text box and press Submit.



Step 4. After doing Submit the software will execute and put the result on the screen



5 Conclusions

In this project, we have done review of various existing techniques and tool for base64 encoding decoding. We have gone through various useful technologies, algorithms and methods for encoding and decoding system.

5.1 Future perspective of Work

- The tool can be made more user friendly by adding more buttons like clear, color etc.
- Other encode decode feature can be added like zip/unzip and other things.
- Python other platform can be used for increasing speed of decode.

- The tool can be developed to run other platform like Linux and Mac os.

References

- [1] Somchai, Wen Dong , “Research on Base64 Encoding Algorithm and PHP Implementation” 2017.
- [2] A. A. Abd El-Aziz, “JSON Encryption” 2014 International Conference on Computer Communication and Informatics 2014.
- [3] Karthik, Muruganandam Data Encryption and Decryption by Using Triple DES and Performance Analysis of Crypto System. 2014.
- [4] Liuyang, “Research and Realization of Base64 and 3DES Encryption on iOS”, Computer Security 2012.
- [5] Luojianghua, “MD5-Base64 based hybrid encryption algorithm”, Journal of Computer Applications 2012.
- [6] G. Wang. Improving Data Transmission in Web Applications via the Translation between XML and JSON. In Proceedings of 2011 Third International Conference on Communications and Mobile Computing (CMC) 2011.
- [7] COVER T, THOMAS J. Elements of Information Theory [M]. 2nded. Hoboken: Jahn Wiley and Sons, 2010
- [8] Aman Kumar, Sudesh Jakhar, Sunil Maakar, “Distinction between Secret key and Public key Cryptography with existing Glitches”, Volume: 1, 2010.
- [9] Khalid Alnafjan, T. H., Gul Faraz Khan, Hanif Ullah, Abdullah Sharaf Alghamdi (2010). "Comparative Analysis and evaluation of Software Security Testing Techniques.
- [10] Abdullah Sharaf Alghamdi, I. A., Muhammad Nasir (2010). Evaluating ESB for C4I Architecture Framework Using Analytic Hierarchy Process.
- [11] Bruno Gil, P. T. (2010). Impacts of data interchange formats on energy consumption and performance in smart phones
- [12] K. D. Lewis dan J. E. Lewis, “Web Single Sign-On Authentication u Journal of Computer Science Issues, vol. 2, no. 1, pp. 41-48, 22 3 2009.

[13] L. Cantara, METS: The Metadata Encoding and Transmission Standard, vol. 4, Cleveland: The Haworth Press, 2005.

[14] BLUNDO C, D' ARCO P, SANTIS A, et al. Design of self-healing key distribution schemes [J]. Designs, Codes and Cryptography 2004,

[15] K. Solanki, V. Vankani, P. Pukle dan S. Iyer, "Multimedia Encryption Using Visual Cryptography", International Journal of Recent Trends in Engineering & Research 2004.