# A Novel Method for CCSDS-TC Recognition

Shen Wang[1], Fanghui Sun[1], Jeng-Shyang Pan[2,3,4], Hongli Zhang[1]

[1] Department of Computer Science and Technology, Harbin Institute of Technology, China
[2] College of Computer Science and Engineering, Shandong University of Science and Technology, China
[3] Fujian Provincial Key Lab of Big Data Mining and Applications, Fujian University of Technology, China
[4] Department of Information Management, Chaoyang University of Technology, Taiwan
shen.wang@hit.edu.cn, sunfanghui@hit.edu.cn, jspan@cc.kuas.edu.tw, zhanghongli@hit.edu.cn

## Abstract

The continuing development of space network communication has put forward higher requirements to space data communication. At the same time, the complicated CCSDS (Consultative Committee for Space Data Systems) network protocol hosting relationships has made the protocol recognition work become more and more difficult. In this paper, we introduced a recognition model to distinguish CCSDS-TC based on its data features in data link protocol sublayer together with synchronization and channel coding sublayer. In the process of feature recognition, signification degrees and feature weights are applied to improve the recognition accuracy. From the experiment results we can find that the proposed method can distinguish CCSDS-TC from other protocols significantly.

**Keywords:** Protocol recognition, CCSDS-TC, Tire tree, Network security, Information security

## 1 Introduction

The diversification of spacecraft platform has made the increment in the variety of space mission and the functional requirement of space mission become more and more complicated. In the 1980s, the Consultative Committee for Space Data System (CCSDS) was founded to standardize the space data communication mission, to enhance the cooperation and support in space network between countries around the world. After the development of more than 30 years, many countries including USA, Russia, German, France, UK have entered this organization. So far, CCSDS has developed a series of standardized space network communication protocols and made many times of improvement to fit the space mission demands better. According to an incomplete statistic, there are more than 260 space missions has adopted the advice about space network communication protocol provided by CCSDS.

Similar to the OSI Basic Reference Model and relative network configurations on the ground [1, 8], protocol family of CCSDS is divided into 5 layers: Physical Layer, Data Link Layer, Network Layer, Transport Layer and Application layer, as is shown in Figure 1.

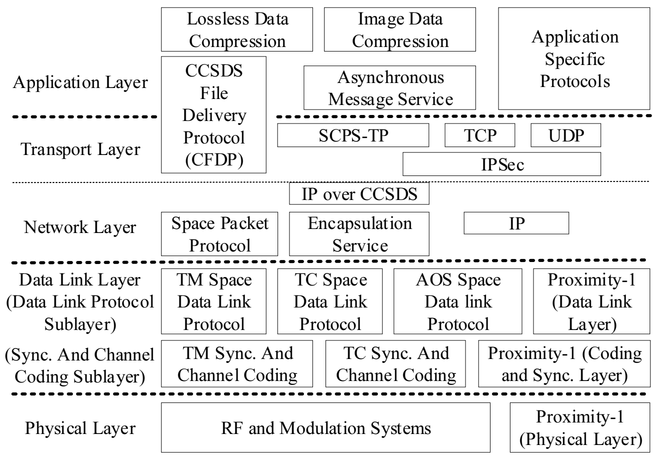


**Figure 1.** Space communications protocols reference model [2]

From the space communications protocol reference model, we can see that there are more than one protocols distributed in each layer. Take the Data Link Protocol Sublayer for example, there are 4 protocols, which are: CCSDS-TM, CCSDS-TC, CCSDS-AOS and Proximity-1. Besides, the hosting relationships between two layers is not unique. For example, the CFDP protocol data unit can be delivered by SCPS-TP in transport layer, and it also can be delivered by SPP in network layer, as is shown in Figure 2.

Compared to the situations in TCP/IP among different layers, the hosting relationships in CCSDS protocols are more flexible in order to satisfy the variant data transport demands in different space missions, which means, different space mission or one space mission at different time can adopt different protocol combinations to implement the data transport.

*Corresponding Author: Jeng-Shyang Pan; E-mail: jspan@cc.kuas.edu.tw
DOI: 10.3966/160792642020032102001

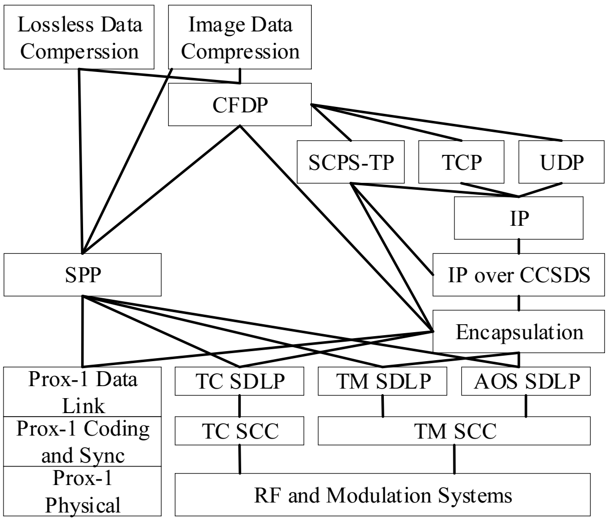


**Figure 2.** Possible combinations of space communications protocols [2]

In this circumstance, techniques about specific space communication protocol recognition is of great important. There are many works about traditional network protocol recognition, especially on application protocols like P2P protocol. As the development on space communication technology, some research works on space network communication protocol (CCSDS protocol family) show up. But most of them are focused on the performance analysis of space communication protocol and the combination between space communication protocol and networks on the ground. There are few works on space communication protocol recognition.

At first, protocols on application layer had specified access ports. So, the port number could be used as a protocol feature. But later, some private application layer protocols use different access port every time when they communicate to avoid being detected. In this way, the access port number could no longer be a protocol feature.

Chen et al [3] raised an idea to recognize application layer protocol based on matching feature strings, which were some certain fields of the target protocol and would show up frequently. This was similar with the method based on port number, but it is more accurate though has a bigger time cost. Zhu [4] put up a protocol identification method based on multi-feature HMM (Hidden Markov Model). He took features from the process of communicating, such as size of TCP packet, timestamp, and so on. Those features cannot be concealed even by encryption. Clustering and quantifying were used in the reprocess stage to increase the recognition performance and received a higher accuracy both on plain and cipher packets.

Although referred to the Internet TCP/IP protocol family layer structure, CCSDS space communication protocol family still has relative bigger delays when transporting, and a higher error rate. Besides, when CCSDS building the advices of space communication protocols, they reduced the number of information fields in protocol format, and many fields were defined to be optional item. Those increased the difficulties to extract protocol format feature.

Yao and Li [5] put forward a recognition method to identify CCSDS protocols in data link layer. By analyzing four protocols in data link layer and building feature models (such as Sync Header model, Pseudo-stochastic model, frame length model and so on), they identified protocols by feature matching. In the work of Li et al [6], they took the starting sequence in remote control channel transmission unit as the protocol feature, and using pattern matching to recognize the CCSDS-TC protocol. But they just used traditional pattern matching algorithms to match the start code and did not took the feature of protocol data in to consideration. Later, Li et al. [7] put up a refinement algorithm based on fast search algorithm to detect the sync word of CCSDS protocol. That algorithm made up the deficiency of classical matching algorithm and had a relatively higher performance.

In this paper, we focused on the CCSDS-TC protocol. After analyzing the specification of CCSDS-TC thoroughly, we found several important features and designed a CCSDS-TC protocol recognition system further. From the experiment results, this system showed a satisfying performance and could recognize CCSDS-TC accurately and efficiently.

This paper is organized as bellow: section 1 talks about the background and related work, section 2 introduces the outline about CCSDS-TC and analyses recognizable features of it, section 3 gives the details of the proposed CCSDS-TC protocol recognition system, the related experiments and result analyses are shown in section 4, section 5 makes a conclusion of this paper.

## 2 CCSDS-TC Introduction and Feature Analysis

### 2.1 CCSDS-TC Introduction

With the continues development of space network communication technology, the traditional measurement and control technology can no longer fit for new space mission which are calling for more effective solutions. In recent years, many countries select new recommended standard proposed by CCSDS and obsolete old techniques gradually.

CCSDS-TC is a link control protocol of the space in data link layer, in charge of transmitting remote control commands and data. This protocol is designed to satisfy the requirements of transmitting various types and properties of space application data in the communication link of ground to space or space to space. Figure 3 shows the CCSDS-TC protocol’s relationship with OSI layer model and CCSDS layer model. The data link layer is divided into two sublayers to describe CCSDS space data link protocol. The TC space data link protocol is located at the data link protocol sublayer, in charging of transmitting data by variable sized data unit. The TC channel coding and synchronization protocol is located at the synchronization and channel coding sublayer, in charging of performances in communication such as error-correction coding, decoding, synchronism and so on.

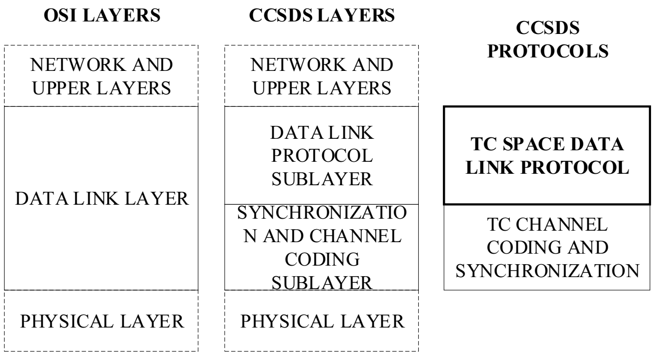


**Figure 3.** Relationship with OSI layers and CCSDS layers [9]

By sending remote control command or data through CCSDS-TC, this protocol is mainly used in control links between earth and satellite or satellite and satellite. According to the protocol standard, the data need to be assembled in turn by packaging layer, segmentation layer, transport layer, channel coding layer, and at last modulated and sent by high frequency carrier signal in physical layer. As it is shown in Figure 4, the ground remote control source packages the data needed to be sent to satellite in the form of remote control package. Then, the package is segmented into one or more remote control segments in the segmentation layer. The transport layer packs the segments into transfer frames and adds heading information in to each frame. The channel coding layer is in charge of data encoding. After the data is sent to physical layer, it will be sent by carrier wave. Correspondingly, the satellite system will parse the data in the turn of physical layer, coding layer, transport layer, segmentation layer, packaging layer, and finally get the data sent by ground remote control source.

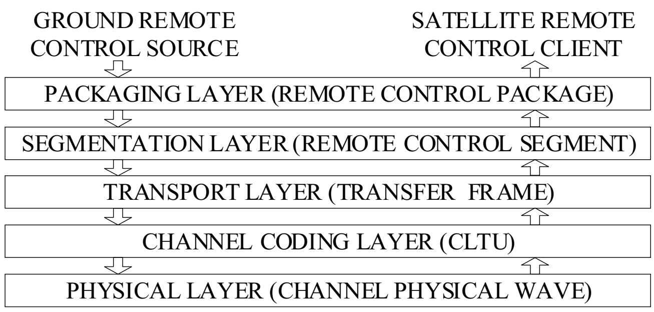


**Figure 4.** System layer structure [9]

From the initial remote control package, to the final physical carrier wave, the format of package is changed when passed through each layer. In one transport mission, there might be one or many packages, with each package has a length of 1~65536. To make the transmit efficiency enough, and make it easy in package retransmitting, the remote package needs to be divided into relative small data segments. The length of a remote control segment is 251 bytes at most, made up by one byte's header and a data field which is no more than 250 bytes. After that, the remote control segment will be added to a frame header which is 5 bytes in transport layer. At this point, one initial remote control package might be divided into one or more transfer frames, and these frames will make up a frame sequence. In the channel coding layer, each frame sequence is encoded by BCH code and become a BCH code block. Then the Communications Link Transmission Unit (CLTU) will be formed by combining all the code blocks, adding a start code ahead and an end code behind. At last, CLTUs will be sent by high frequency carrier wave. During all the transmission process, the packaging sequence of data package format is shown in Figure 5.

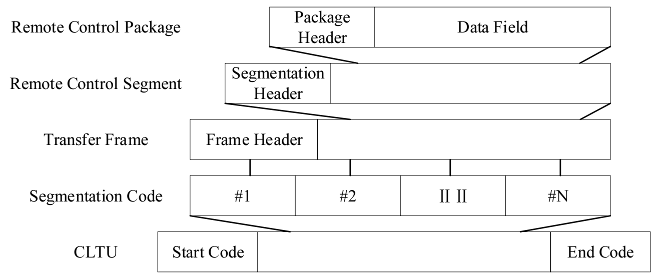


**Figure 5.** Segmentation process [9]

## 2.2 CCSDS-TC Feature Analysis

This system focuses on the protocol format to identify CCSDS-TC. In the aforementioned five layers processing of CCSDS-TC, the packaging layer and segmentation layer is mainly used to package the remote control data, and they are not closely related to the protocol's inherent feature. So, the recognizing system concentrates on data structures in sthe transport layer, synchronization and channel coding sublayer, and physical layer, to extract features.

A transfer frame is mainly composed of three parts: Transfer Frame Primary Header, Transfer Frame Data Field, and Frame Error Control Field. Among them, the Header includes some important information about the transfer frame, like Frame Length, Sequence Number, Spacecraft ID, Virtual Channel ID and so on. The data in the Transfer Frame Data Field comes from the remote control package in upper layer. The Error Control Field is optional. The concrete format of a TC Transfer Frame is shown in Figure 6. The length of a transfer frame is changeable, and at most is 1024 bytes. Inside, the Primary Header is pretty important, the detailed structure is shown in Figure 7. the Frame Length field is 10 bytes, which means it can indicate at most 1024 bytes, and this is consistent with the biggest length of the transfer frame.

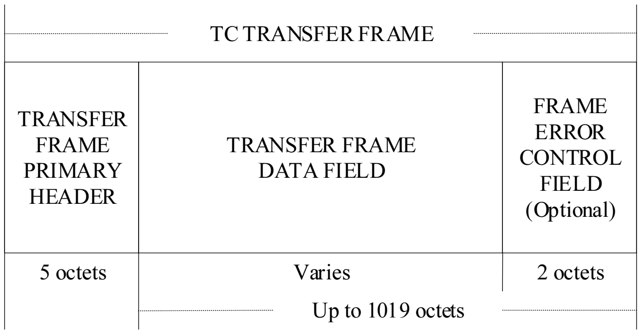


**Figure 6.** TC transfer frame structural components [9]

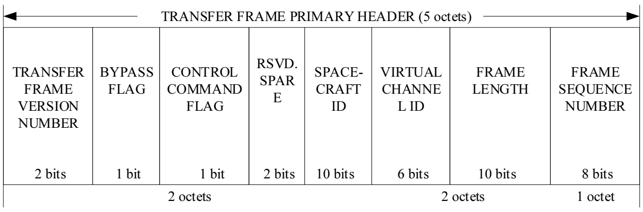


**Figure 7.** Transfer frame primary header [9]

The format of CLTU is shown in Figure 8. There are 3 parts in CLTU: Start Code, BCH Code Block Stream, and End Code. The remote control code block stream is the set of blocks obtained by BCH encoding the transfer frame sequence. Each of the block is 8 octets. The Start Code of CLTU stands for the beginning of remote control encoded data and the operation of decode. The Start Code has 16 bits and are set to be EB90 (in hexadecimal format) in CCSDS-TC recommended protocol standard. The End Code stands for the end of CLTU and the operation of decode. It has a length of 64 bits and are designed to be C5C5C5C5C5C5C579 in CCSDS-TC.

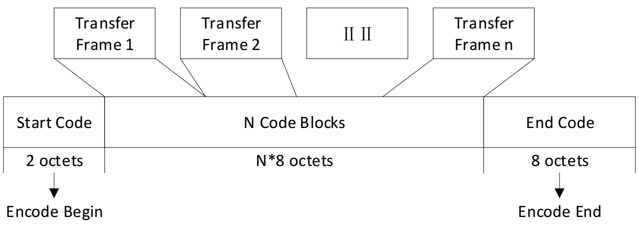


**Figure 8**. CLTU format [9]

There are also standard data structure in physical layer, which is composed of Starting Sequence, CLTUs, and Idle Sequence, as is shown in Figure 9. The physical layer is in charge of modulating and using high frequency carrier waves to send out the CLTUs received from the synchronization and channel coding sublayer. A physical link will be established between the sending end and the receiving end before the physical layer transmitting data. After a series CLTUs arriving the receiving end, the communication conversation will be ended by the vanishing of the carrier waves. The Starting Sequence is made up by alternating '0' and '1'. Its length is determined by the performance requirement of communication link but is no shorter than 16 octets. The starting sequence is aimed to obtain the initiated bit synchronization. CLTU is the above-mentioned data structure in the synchronization and channel coding sublayer. The idle sequence is used to maintain the bit synchronization when there are no CLTUs. It's made up by alternating '0' and '1', and its length is variable.

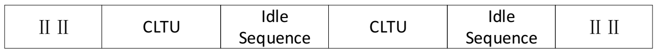


**Figure 9.** Data structure in physical layer

## 3 System Construction and Realization

After analyzing the process of communication and packaging, several features were figured out from data link sublayer, Synchronization and channel coding sublayer and physical layer in this section. In our CCSDS-TC protocol recognition system, we defined significance degrees for the features we generalized and set specified weights for them. During the process of data analyzing to identify CCSDS-TC, a revised feature matching algorithm will be used to measure whether the data were produced by the target protocol.

### 3.1 System Structure

The structure and data analyzing process is shown in Figure 10. As is mentioned above, there are several features in CCSDS-TC protocol, such as the frequency appearance of idle sequence, the remaining of the spacecraft ID in the Header of transfer frame. But these features are distributed in different protocol levels. The idle sequence is frequency showed in physical layer, and it is in data link layer that the spacecraft ID in frame header keeps the same. We cannot confirm a data sequence to be CCSDS-TC just because of the matching of one feature sequence but should judge it according to all the features' matching states. In the recognizing process of CCSDS-TC, the slitting of data is based on one or more features of CCSDS-TC, and the results of data slitting have their own data features. In this paper, we abstracted five features according to the data formats of CCSDS-TC in physical layer, synchronization and channel coding sublayer, and data link sublayer, and defined corresponding significant levels ($s_i$) for each feature to indicating how sensible the features are. We also assigned different weights ($w_i$) for each feature's significant level. The final recognition of CCSDS-TC is decided by the sum of each feature's significant level times its weight.

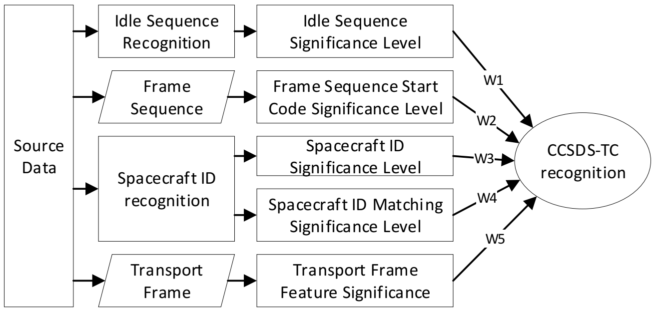


**Figure 10.** CCSDS-TC recognition model

## 3.2 Features Recognition and Their Weights

### 3.2.1 Significant Level of Idle Sequence (SLIS)

For the frequently appeared idle sequences, they also follow certain rules for their locations in the protocol data. The idle sequence takes one '10101010' as a unit. It usually appears sparsely after shows up some times continually, and then comes out some times again, and so goes on in a circle. Here, we defined the count of idle sequence continual appearance. And the continual appearance means the difference value of two idle sequence units' location is the length of one unit. We take '10101010' for example. When a bit sequence '1010101010101010' appears in a data stream, we consider that the idle sequence unit continually appeared once. The location list and real appearance time can be calculated by AC automation matching. According to the location list, we can figure out the total times of idle sequence continually appeared. The significant level is defined as the ratio of the total times of continually appeared and the real times of appearance. This value should be the real times of showing up subtracts the numbers of blocks divided by real times of showing up. But, as the idle sequence is an 8 bits string, there might be many single '10101010' strings in a big amount of data. These single unit strings directly influenced the measure of continual appearance of idle sequence, thus we need to avoid this impact. Set N to be the real times of idle sequence showed up, n to be the time of continual appearance. In the process of calculating n, if there is single appeared '10101010' string, we remove the single string's influence by making N=N-1. At last, the continue degree is defined as the ratio of times of continual appearance and the times of real appearance, which is n/N. Here, the N is no longer the one got by AC automation, but much smaller than that.

### 3.2.2 Significant Level of Frame Sequence Start Code (SLFSSC)

The data structure of a frame sequence is introduced in last section. A frame sequence is also a channel transform unit, which is also called CLTU. An entire frame sequence includes three parts: start code, BCH code block, and end code. It is mentioned in the CCSDS-TC standard recommendation that, the start code and end code in one integral transport mission keep the same, which means that in each frame sequence, the preceding 16 bits and the following 64 bits are consistent theoretically. That is one feature of CCSDS-TC in synchronization and channel coding sublayer.

Theoretically, the preceding 16 bits in all the frame sequences should be the same. But, there might be error or missing in the process of cutting the initial data to frame sequence. Or even worse, if the idle sequences are error recognized, the resulting frame sequences will be totally wrong. And there might be singular values in the preceding 16 bits in the frame sequences or all the values are distributed. Therefore, we need to measure the significance level of the start code of frame sequence to assist the recognition of CCSDS-TC protocol.

The way to measure the significance level of the start code of frame sequence is as bellow. Firstly, take the frame sequence as the object, and statistic the values of the preceding 16 bits $S(s_1, s_2, \ldots, s_n)$. Theoretically, the values in S should be unanimous, or exists several singular values occasionally. Set $N_{mode}$ to be the number of the mode of S, and the significance level of frame sequence start code is define as the ratio of $N_{mode}$ and n.

### 3.2.3 Significant Level of Spacecraft ID (SLSI)

The spacecraft ID can be obtained by calculating the frequency items, and we will explain it in next section. In the process of analyzing frequency items, we take a frame sequence as a unit. As the number of frame sequence divided from the initial data might be great large, and the frame sequence data is storage in the form of text, it may result in the decreasing of the performance of spacecraft ID recognizing program. Therefore, in this paper, we calculate the frequency item from the samplings of frame sequences to increase the efficient of the program and reduce the probability of spacecraft ID error recognition caused by incomplete frame sequence statistic. After the spacecraft ID is decided, we take the ratio of number of spacecraft ID and number of frame sequence sample to be the signification of spacecraft ID to assist the recognition of CCSDS-TC protocol.

### 3.2.4 Significant Level of Spacecraft ID Matching (SLSIM)

The process of dividing the frame sequence into frames is based on the recognition of spacecraft ID, and the identification of spacecraft ID is by statistic of frequency items. Although the judging of 2 bits of '00' are added into the recognition of spacecraft ID, there might be mistakes in the process of dividing frame sequence into frames. Thus, we defined spacecraft ID matching significance level to assist the recognition of

CCSDS-TC protocol.

By analyzing the header of transfer frame, we can find that, 2 bits in the front of the frame is the transfer frame version number, which are set to be constant ‘00’ to differ from other data link sublayer protocols like CCSDS-TM, CCSDS-AOS, and others. Besides, there are another 2 bits to be constant ‘00’ after the virtual channel ID field. These two features are the inherent characters of CCSDS-TC in data link sublayer. The significant level of spacecraft ID matching is defined based on them.

The concrete way to calculate is as bellow. In the process of dividing frame sequence into frames, when matching the spacecraft ID, set it as the basis, check the corresponding version field and virtual channel ID field to see if the certain places are ‘00’ according to the CCSDS-TC protocol standard. Plus 1 if match, keep the same otherwise. When finishing the dividing process, the ratio of matching count and numbers of spacecraft ID will be the significant level of spacecraft ID matching.

#### 3.2.5 Significant Level of Frame Format Feature (SLFFF)

The potential mistakes of recognizing spacecraft ID in the process of dividing frame sequence might cause errors in the later data field dividing step. So, the significant level of frame format feature is defined to assist the recognition of CCSDS-TC.

According to the definition of fields in the header of CCSDS-TC transfer frame, the length field represents for the length of the frame counted by bytes, which means that the value of the length field is equal to the real length of the transfer frame. This can also be treated as a feature of CCSDS-TC in data link sublayer and can also be a reference of measure the significance level of frame format feature. As the process of dividing frame sequence into frames is based on the spacecraft ID, so, if there is bit sequence in data field which is same with the spacecraft ID, the division would be wrong. This is only a special condition, but if the data being divided is not the standard CCSDS-TC protocol package, there will be massive error divisions.

Here we present a way to measure the significance level of frame format features. For each frame being divided into, parse and recognize it according to the standard CCSDS-TC protocol transfer frame format. If the value of a frame length field is equal to the whole length of that frame, then the frame format feature matching value would be added by 1. Keep it the same otherwise. At last, the ratio of frame format feature matching value and the number of values will be the wanted significance level of frame format feature.

#### 3.2.6 Weights of significance level values

In the CCSDS-TC recognition model, we set different weights for each significance level of protocol feature. The rationality of distributing the values of weights directly effects the accuracy of the final CCSDS-TC recognition. So, in this paper, we adopted a principle of increasing allocation to set the weight values. That is, the more fine grained the data division is, the higher the relevant weight value of a feature’s significance level will be. For the five significance levels of CCSDS-TC protocol features, their weights are allocated to be 0.1, 0.15, 0.15, 0.2, 0.4. The sum of these weights is 1. Then we explain the allocation reasons bellow.

The initial data is divided twice from the physical layer to synchronization and channel coding sublayer to data link sublayer. For the initial data, we don’t know which kind of protocol it is, and the whole process of CCSDS-TC protocol recognition is based on the assumption that the data is belong to CCSDS-TC protocol. We use the idle sequence to divide the data from initial data to frame sequences. But the existence of idle sequence cannot indicate that the data is from CCSDS-TC protocol. From the survey of other protocols in CCSDS we know that CCSDS-TM protocol also has idle sequence in physical layer. Suppose the initial data is CCSDS-TM, the value of the significance level of idle sequence should also be pretty high. Therefore, we assign the lowest weight for the feature of idle sequence. The division from frame sequence to frames is based on the spacecraft ID, which is one of the features of CCSDS-TC transfer frame. From the research of other protocols in CCSDS data link sublayer we found that the transfer frame of CCSDS-TC is very different from that of CCSDS-TM. So, we assign relative high weights for the later three significance levels. This kind of weight allocation can reduce the false positive rate of the identification of CCSDS-TC caused by the similarity of protocol formats.

### 3.3 Feature Sequence Recognition

The five features and their significance levels and weights are introduced in last subsection. In the process of dividing data, the recognition of idle sequence and spacecraft ID are needed to support other analysis. In this paper, we adopt the techniques of frequency item mining to handle this problem.

#### 3.3.1 Idle Sequence Recognition Based on Frequency Item Mining

The main features of CCSDS-TC are centered on the transport layer, syntax and coding layer, and physical layer. Each layer has its own features. The protocol recognition method of pattern matching based on only features of only one layer is not satisfying both in the recognition aspect and the efficiency aspect. For example, in the synchronization and coding sublayer, the data structure of protocol is CLTU. According to the introduction in 2.2, there is a start code in the CLTU format which is setup to be 0xEB90. In the CCSDS protocol family, there is another protocol in the synchronization and coding sublayer called CCSDS-TM, and it has a start code of 4 bytes. So theoretically, the CCSDS-TC and CCSDS-TM can be identified just by pattern matching. However, experiments indicated that this method can identify CCSDS-TC to some extent, but with very low accuracy. The reason is that, in the whole protocol data, the start code is a small part, and it is random appeared, which including some cases that bits in data field has the same bits with start code. To execute the subsequent dividing operations accurately, we need more other features to divide the initial data into frame sequence. Before the protocol recognition process, the data from physical might not be complete. But according to the protocol standard showed in Figure 9, the idle sequence and CLTU should take place by turn. The official recommendation of idle sequence from CCSDS is like this: The Idle Sequence is the data structure which provides for maintenance of symbol synchronization in the absence of CLTUs. The bit pattern is a sequence of alternating 'ones' and 'zeros'. The length of the Idle Sequence is an unconstrained number of bits [10]. According to this description, we find that there is a regular pattern in idle sequence that the '0s' and '1s' take turns to show up, and the idle sequence take 8 bits as a unit. The relative high appearance and showing up every 8 bits are all the features of the idle sequence. Therefore, we take the idle sequence to divide the initial data into frame sequences.

According to the above features of idle sequence, we proposed an idle sequence recognition algorithm, which including two steps:

(A) Frequency Item Statistic. Calculation of frequency items based on the Tire Tree method and obtain N candidate frequency item sets.

(B) Frequency Item Filtration. Filter the N frequency item sets according to the following strategy and confirm one to be the idle sequence. Divide the initial data into frame sequence according to the idle sequence.

The purpose of fliting the frequency item is to find the one which fits the idle sequence features most. There are three steps.

Step one: rank the frequency items according to their appearing times, and then select the former n items to make the set for next selection by calculating the range of variation of two adjacent items.

Step two: construct AC automation by n frequency items, which are also n pattern strings. Iterate the initial data to get the location sequence of the n frequency items.

Step three: for each frequency item in n frequency item sets, compute the continuity degree according to the difference of the front and back positions in their location sequence. Take an 8-bits frequency sequence for example, if the difference of two continuous appearing is equal to 8, then the continuity degree should be added by 1. Keep it the same otherwise. In the last statistic, the item having the highest continuity degree will be considered to be the idle sequence.

### 3.3.2 Spacecraft ID Recognition Based on Frequency Item Mining

By the division from the initial data, we can obtain a series of frame sequences data with each frame sequence stored separately for the following data division work. According to the recommendation of CCSDS-TC in synchronization and coding sublayer, the number of frame in frame sequence is uncertainty. There might be one or more frames in a frame sequence. As the recognition process of division from the initial data to frame sequence is based on the idle sequence obtained from frequency item statistic, and that does not make sure the complete accuracy of the division, so we cannot divide the frame sequence into frames simply based on the transfer frame format.

The format of transfer frame is as shown in Figure 6 and Figure 7. From the analysis of transfer frame, we can find there is a spacecraft ID field which is 10 bits. According to the recommendation of CCSDS-TC, the Spacecraft Identifier is assigned by the CCSDS and shall provide the identification of the spacecraft associated with the data contained in the Transfer Frame. The Spacecraft Identifier shall be static throughout all Mission Phases [7]. Theoretically, in the transfer frame sequences obtained by division based on idle sequence, the value of the spacecraft ID field in every transfer frame header is settled, which is an important feature in the frame sequence data. Therefore, in this paper, we take the spacecraft ID as basis to divide the frame sequence into separated frames.

Similar with the idle sequence, in the recognition process from frame sequence to frame, the processed data is still bit stream data. So, the confirmation of spacecraft ID is similar with idle sequence, that is the method of frequency item statistic but just has a little difference about the way to filter the frequency items. The concrete recognition procedure is as bellow.

Step one: frequency item statistic. By checking the format of transfer frame primary header in Figure 7 we can find that, the spacecraft ID field takes 10 bits and there is a 2-bit idle field assigned to be '00' before the spacecraft ID field. So, when recognizing the spacecraft ID, we take this field and the idle field together into consideration and calculate the 12-bit frequency item. By statistic the frequency item in every frame sequence, we can find each 12-bit frequency item for every frame sequence.

Step two: the confirmation of spacecraft ID. As the 12-bit frequency item in each frame sequence can be not only one, so it is needed to filter the frequency items to confirm the real spacecraft ID. We make twice filter for the frame sequence here.

(a) as the statistic of frequency item is based on the 12 bits of spacecraft ID field and the former 2-bit idle field, so we can filter the frequency items once by checking if the frontal 2 bits of the frequency item is ‘00’.

(b) as the frame sequence is made up of complete frames, the location of the frequency item we want to find should be at the front of the frame. But in fact, there might be error judges. So here, we take the location one frequency item first shows up and its frequency count to be its features and assign different weights for them. The allocation of weights is lean to the location it first appears, and the difference of two features times each one’s weights will be checked. The minimum difference value belongs to the real spacecraft ID field of the present frame sequence.

According to the former two steps, a 12-bit frequency item is obtained for each frame sequence data. After statistic of all frequency items, the one with the highest frequency count will be recognized as the spacecraft ID and will be used to divide the frame sequence into frame data.

## 4 Experiments and Analysis

The CCSDS-TC protocol recognition method proposed in this paper identify the unknown protocol under the assumption that the target protocol might be CCSDS-TC, and then analysis the data according to the features of CCSDS-TC. In the process of recognition, five feature significance levels in the CCSDS-TC recognition model are calculated, and at last the judgement is made based on the output of this model. Consider about the specialty of space network communication, the initial data is considered to be singular protocol in this paper other than hybrid data.

In this paper, the threshold of the result is assigned to be 0.8. If the recognition result is larger than 0.8, then the recognized data is considered to be of CCSDS-TC, vice versa.

The experiment is taken for several different types of protocol, and the data of each is simulated according to its standard format. For each protocol, 10 groups of data are generated and used to test our CCSDS-TC recognition model.

### 4.1 Simulation Data Test for CCSDS-TC

By the simulation program designed by the official recommendation of CCSDS-TC, the data for test data is generated and tested. The recognition result is shown in Table 1. From the recognition results we can see that all the recognition results of CCSDS-TC protocol are above 0.85. And in the 10 experiment tests, the recognized frame sequence number is same that when the data is simulated, which indicates the correctness of idle sequence recognition and the division from initial data into frame sequences. Besides, the difference between frame number and the real number of frame when simulated is less in 10. By debugging, the error is caused by the spacecraft ID sequence appeared in the data field which leads to the mistaken in frame division.

### 4.2 Simulation Data Test for CCSDS-TM

The simulation data of CCSDS-TM is generated according to the official recommendation. Consider that the packet format of CCSDS-TM is also frame sequence in synchronization and coding sublayer, the experiment of CCSDS-TM protocol data can be used to test the accuracy of idle sequence recognition in our model. The recognition result is shown in Table 2.

**Table 1.** CCSDS-TC recognition result

| Number | SLIS | SLFSSC | SLSI | SLSIM | SLFFF | Result |
|---|---|---|---|---|---|---|
| 176 | 0.97 | 0.99 | 0.92 | 0.68 | 0.92 | 0.89 |
| 56 | 0.96 | 0.98 | 0.85 | 0.69 | 0.92 | 0.88 |
| 179 | 0.96 | 0.99 | 0.96 | 0.69 | 0.92 | 0.90 |
| 69 | 0.95 | 0.98 | 0.88 | 0.69 | 0.92 | 0.88 |
| 165 | 0.96 | 0.99 | 0.92 | 0.69 | 0.93 | 0.90 |
| 192 | 0.96 | 0.99 | 0.94 | 0.68 | 0.92 | 0.90 |
| 121 | 0.96 | 0.99 | 0.92 | 0.68 | 0.92 | 0.89 |
| 179 | 0.96 | 0.99 | 0.9 | 0.69 | 0.92 | 0.89 |
| 180 | 0.96 | 0.99 | 0.9 | 0.69 | 0.92 | 0.89 |
| 109 | 0.96 | 0.99 | 0.92 | 0.69 | 0.91 | 0.89 |

**Table 2.** CCSDS-TC recognition result for CCSDS-TM data

| Number of Frame Sequence | Recognition Result |
|---|---|
| 167 | 0.04 |
| 146 | 0.04 |
| 115 | 0.04 |
| 193 | 0.05 |
| 152 | 0.05 |
| 196 | 0.04 |
| 155 | 0.04 |
| 107 | 0.04 |
| 156 | 0.04 |
| 133 | 0.04 |

From the 10 data in Table 2 we can find that all the result is nearly 0. In the experiment process, for the five protocol feature significance levels, only the one of idle sequence is 0.4, and all the others are 0. That was because no frame sequence was obtained in the division from initial data to frame sequence, which made the program cannot go on the subsequent operations and the rest features cannot be found. In physical layer, the CCSDS-TM is made up of alternating of start code and encoded transfer frame sequence. There is no idle sequence in consecutive frame sequences. Besides, the format of CCSDS-TM protocol transfer frame is quite different with CCSDS-TC. Although there is spacecraft ID field in CCSDS-

TM transfer frame header, the absence of idle sequence feature in physical layer makes the result to be almost 0. The recognition result tells that the object data is not CCSDS-TC protocol data.

### 4.3 Simulation Data Test for GJB-PCM

The PCM data is simulated according to the national standard of PCM communication protocol to test the efficiency of this CCSDS-TC recognition model as the packet format of PCM is similar with CCSDS-TC. The recognition results are shown in Table 3.

**Table 3**. CCSDS-TC recognition result for PCM data

| Data Size (KB) | Recognition Results |
|---|---|
| 719 | 0.37 |
| 595 | 0.39 |
| 606 | 0.37 |
| 733 | 0.38 |
| 521 | 0.37 |
| 667 | 0.37 |
| 628 | 0.37 |
| 665 | 0.37 |
| 689 | 0.52 |
| 528 | 0.36 |

The PCM remote control is a traditional space telecommunication technique. There are a lot of similarities of CCSDS-TC and PCM protocol in the aspects of packet format, remote control protocol layering, and communication process. Although it has the features of frame sequence, transfer frame, and start code, the recognition result for the 10 groups of PCM remote control data is almost bellow 0.4. By observing the five values of protocol feature significance level we found that the significance level of idle sequence and frame sequence start code are both around 0.98. That is because the similarity of PCM and CCSDS-TC in physical layer packet format which is the alternation of ide sequence and encoded frame sequence. So, the program has a good accuracy in the recognition of idle sequence and the division from initial data to frame sequences. The value of spacecraft ID is about 0.8. That is because of the functional similarity of the satellite address synchronization code in PCM protocol transfer frame and the spacecraft ID field in CCSDS-TC transfer frame header. Both of them can be seen as the address of a spacecraft, and it remains the same in the transport mission. The significance level of spacecraft ID matching and frame format are both close to 0. This situation is because of the difference in transfer frame format in data link sublayer, even they have some similarities in physical layer and synchronization and channel code sublayer.

### 4.4 Simulation Data Test for Other Data

At last the CCDS-TC recognition model is tested by the random data. The result is as shown in Table 4.

**Table 4.** CCSDS-TC recognition result for random data

| Data Length (MB) | Recognition Result |
|---|---|
| 64 | 0.05 |
| 24 | 0.05 |
| 40 | 0.05 |
| 72 | 0.05 |
| 32 | 0.05 |
| 24 | 0.05 |
| 72 | 0.05 |
| 56 | 0.05 |
| 48 | 0.05 |
| 16 | 0.05 |

As expected, the recognition result of random data is basically 0. The significance levels are recorded. Only the one for idle sequence is about 0.5. the number of frame sequence divided from the initial data is 0. So the program cannot continue the next recognition operations and the only the significance level of idle can be obtained.

In this section we tested our CCSDS-TC recognition model by four types of simulated data. In the results, the recognition result of CCSDS-TC data is obviously higher than the others. From the result data, we can consider that the recognition accuracy of CCSDS-TC is 100%.

## 5 Conclusion

In this paper, after the analyzing of the communication process and protocol packet format feature, we proposed a CCSDS-TC recognition model by measuring the significance level of its features. The CCSDS-TC protocol is a recommendation standard raised by Consultative Committee for Space Data Systems. The recognition method proposed in this paper is a reversal process of the regular CCSDS-TC protocol data process. According to the data packaging format in the CCSDS-TC communication, we analysis the data layer by layer bottom up and calculate five significance level of protocol features in this process. At last the recognition result given by the recognition model shows of satisfying accuracy.

## Acknowledgements

This work is supported by the National Natural Science Foundation of China (Grant Number: 61471141, 61361166006, 61301099), Basic Research Project of Shenzhen, China (grant Number: JCYJ20150513151706561), and National Defense Basic Scientific Research Program of China (JCKY2016203C004).

## References


[1] J. S. Pan, L. Kong, T. W. Sung, P. W. Tsai, V. Snasel, α-Fraction First Strategy for Hierarchical Model in Wireless Sensor Networks, *Journal of Internet Technology*, Vol. 19, No. 6, pp. 1717-1726, November, 2018

[2] CCSDS Secretariat, *Overview of Space Communications Protocols*, CCSDS 130.0-G-2, December, 2007.

[3] L. Chen, J. Gong, X. Xu, A Survey of Application-Level Protocol Identification Algorithm, *Computer Science*, Vol. 34, No. 7, pp. 73-75, July, 2007.

[4] S. Y. Zhu, *The Study on Protocol Identification Technology*, Ph.D. Thesis, National University of Defense Technology, Changsha, China, 2008.

[5] X. J. Yao, X. Li, Protocol Recognition Technology for the CCSDS Space Data Link Layer, *Aerospace Electronic Warfare*, Vol. 28, No. 2, pp. 26-28, 48, February, 2012.

[6] X. Y. Li, X. Meng, X. J. Yao, Study of Blind Recognition for CCSDS-TC Transfer Frame, *Journal of University of Electronic Science and Technology of China*, Vol. 41, No.1, pp. 142-146, January, 2012.

[7] X. Li, X. J. Yao, X. Y. Li, Improved ASM Detecting Algorithm for CCSDS Protocol Identification, *Avionics Technology*, Vol. 43, No. 2, pp. 36-45, February, 2012.

[8] J. S. Pan, C. Y. Lee, A. Sghaier, M. Zeghid, J. Xie, Novel Systolization of Subquadratic Space Complexity Multipliers Based on Toeplitz Matrix-Vector Product Approach, *IEEE Transactions on Very Large Scale Integration Systems*, Vol. 27, No. 7, pp. 1614-1622, July, 2019.

[9] CCSDS Secretariat, *TC Space Data Link Protocol*, CCSDS 232.0-B-2, September, 2010.

[10] CCSDS Secretariat, *TC Synchronization and Channel Coding*, CCSDS 231.0-B-2, July, 2010.


## Biographies

**Shen Wang** received the Ph.D. degree in computer science and technology from Harbin Institute of Technology, Harbin, China, in 2012. He is currently an Associate Professor of Harbin Institute of Technology, Harbin. His research interests include network security, multimedia security, artificial intelligence and quantum computing.

**Fanghui Sun** is currently working toward the Ph.D. degree in the Computer Science and Technology, Harbin Institute of Technology, Harbin, China. Her research interests include network protocol reverse engineering and network information security.

**Jeng-Shyang Pan** received the Ph.D. degree in Electrical Engineering from the University of Edinburgh, U.K. in 1996. Currently, he is an IET Fellow and Professor in Shandong University of Science and Technology, Fujian University of Technology, and Chaoyang University of Technology. His current research interests include soft computing, information security, big data mining and signal processing.

**Hongli Zhang** received the Ph.D. degree in Computer Architecture from Harbin Institute of Technology, Harbin, China, in 1999. She is currently a Professor and deputy director of Research Center of Computer Network and Information Security Technology, Harbin Institute of Technology. Her research interests include network and information security, network measurement & simulation and grid computing.