Packet Steganography Using IP ID

Asst. Prof. Dr. Ziyad Tariq Mustafa*[1], Authman Waleed Khalid*[2]

*[1] Head of Computer Science Department, Collage of Science, University of Diyala, Iraq

[2]* Computer Eng., University of Technology, Baghdad, Iraq

Received 21 January 2014 ; Accepted 9 February 2014

Abstract

Packet steganography is a secret mechanism that can be used to leak significant information across a network in a manner that violates security policy and it can be difficult to detect. The huge amount of data and vast number of different protocols in the internet seems ideal as a cover for secret messages. In this research, a proposed method is suggested for packet steganography system. The proposed method uses the TCP/IP protocol header field to cover secret message. The secret characters are encoded to ASCII code before merged in IP ID field using specially designed embedding / extracting algorithms in order to make the system more complex to be defeated by attackers. Results are tested successfully with secret message of (42) character, and cover of (100) raw packets.

Key words: Packet steganography, hiding in IP identification, TCP/IP protocol steganography, Covert channel in IP header, Network steganography.

اخفاء البيانات داخل حزم الشبكة في الحقل IP ID

أ.م.د زياد طارق مصطفى *1 عثمان وليد خالد *2

رئيس قسم علوم الحاسوب / جامعة ديالى *1

مهندس حاسوب / الجامعة التكنولوجية *2

Packet Steganography Using IP ID

Asst. Prof. Dr. Ziyad Tariq Mustafa[*1], Authman Waleed Khalid[*2]

الملخص

اخفاء البيانات في حزم الشبكة هي عملية سرية تستخدم لنقل بيانات مهمة خلال الشبكة و من الصعب اكتشافها. ان وجود كمية هائلة من البيانات المنتقلة في الشبكة وعدد كبير من البروتوكولات يوفر بيئة ملائمة لاخفاء البيانات. في هذا البحث قد اقترح نظام اخفاءالبيانات داخل حزم الشبكة باستغلال حقول ترويسة البروتوكول الخاص بالانترنت وبالاخص في حقل التمييز لاخفاء رسالة نصية سرية. يتم تحويل النص الى ما يكافئه من قيمة رقمية متمثلة بالشفرة الامريكية المقياسية وبعد ذلك تقوم خوارزمية التضمين في الجهة المرسلة باخفاء هذه القيمة في حقل التمييز التابع لبروتوكول النقل في الانترنت. وعند الاستلام تقوم خوارزمية الاستخراج بعملية استرجاعالقيمة السرية من حقل التمييز وتحويلها الى النص المكافئ. النتائج في هذا البحث قد خضعت للفحص والتقييم وتم خفاء رسالة نصية سرية متكونة من (42) حرف حيث تم تضمينها في عينة من حزم الشبكة متكونة من (100) حزمة.

الكلمات المفتاحية: الاخفاء في حزم الشبكة,التضمين داخل حقل التمييز في بروتوكول الانترنت, اخفاء في بروتوكولات الشبكات, تواصل سري باستخدام ترويسة بروتوكول الانترنت, الاخفاء داخل الشبكات.

Introduction

The number of articles related to information hiding, network steganography and their techniques has been increased. It has to be taken into account that packet steagnography techniques can be involved in anything based on protocol, even out of the networking scope [1].

The worldwide network of the Internet is the perfect medium for steganography to occur. Data can be hidden in web pages that pass over the Internet, even more surreptitious and unique way to hide messages would be in the unused fields of the TCP/IP packet headers. The operation of the Internet runs on the Transmission Control Protocol and Internet Protocol (TCP/IP) [2].

A protocol header can serve as a carrier for a steganographic data if a header field can take one of a set of values, each of which appears plausible to passive warden. The warden should not be able to distinguish whether the header was generated by an unmodified protocol stack or by a steganographic mechanism [3].

This work is about “Packet Steganography using IP ID field” exploiting TCP/IP traffic to formulate a covert channel transporting secret message.

1.1 TCP/IP Packet Steganography

TCP/IP packet steganography exploits the fact that few headers fields are altered in transit. IP packets can be fragmented, but (unless data are hidden in the fragmentation-related fields) no information is lost. Many header fields can be exploited according to particular packet transmitting situation [4].

Figure (1) illustrates the basic TCP/IP headers with fields shown in (italics and underlined) are those that may be used to embed steganographic data [5].

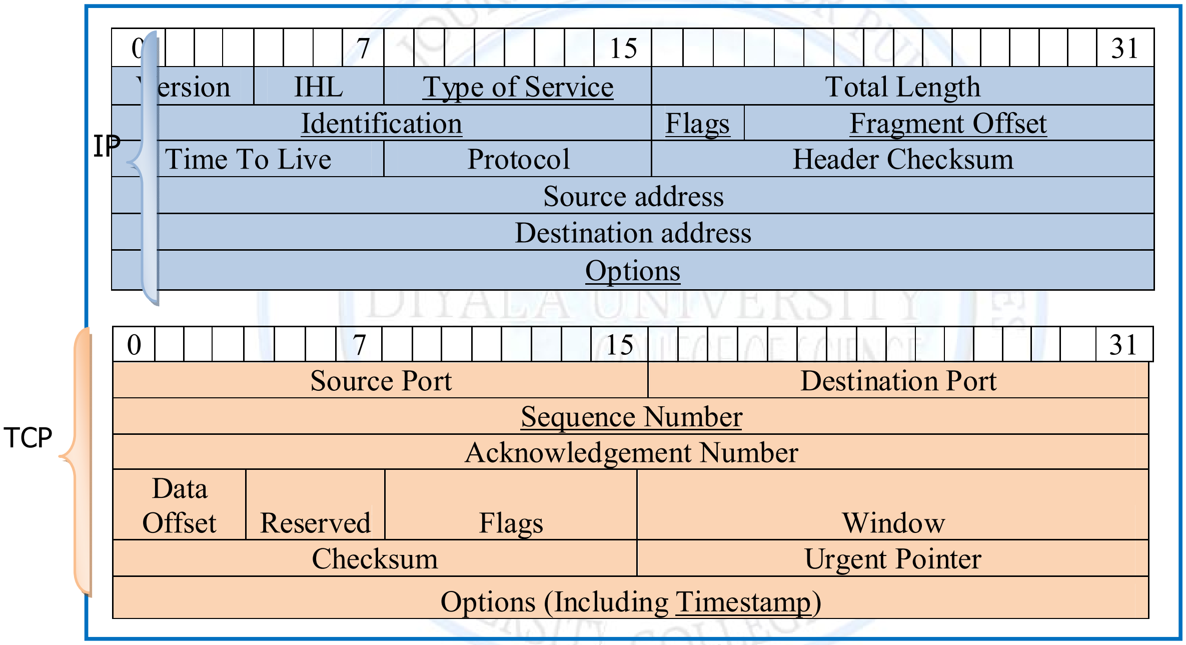

Figure (1) Basic TCP/IP Header Structure after [5]

The fields that are the most interest for use as steganographic cover:

- IP Type of Service: the eight bits of Type of Service (ToS) in the IP header indicate quality of service parameters to routers on a packet’s path. There is potential use of this field as a steganographic carrier, as described by Handel and Sandford [1], because many networks never use them. However, this field would be easily detected by the warden in a threat model, as the field is set to zero in almost all default operating system configurations [4].

- IP Identification: the IP Identification field (IP ID) is “an identifying value assigned by the sender to aid in assembling the fragments of a datagram”, and is allocated 16 bits of the IP header. Because the IP ID is used to distinguish fragments that belong to one packet from fragments belongs to another, it should be unique over the length of time that fragments of a packet might reasonably remain in a network, and unpredictable. [6].

- IP Flags: IP packets include two flags, Do Not Fragment (DF), indicating that the packet should be discarded if it cannot be sent without fragmentation, and More Fragments (MF) which is ‘0’ if the packet contains the last fragment or has not been fragmented. It can be useful to propose the use of the DF bit for steganographic signaling [7].

- IP Fragment offset: when IP packets are fragmented, the individual fragments contain an offset field; this allows the receiving host to reconstruct the fragments in the correct positions in its receive buffers [8].

Information can be transmitted covertly by modulating the size of the fragments originated by a host, and thus the fragment offsets. As with the ToS fields, this method of steganographic encoding could be detected. In environments where path MTU discovery is routinely used, fragmented packets are unusual [9].

- IP Options: the IP packets are contain “options”, so their potential for use in undetectable steganography is limited, some systems describe the use of the IP Timestamp option (not to be confused with the TCP Timestamp discussed later), but in addition to being easily detectable, packets with this option present can travel at most 20 hops, so it is of limited use on the Internet [7].

- TCP Sequence Number: when a connection is established, both hosts must choose an initial sequence number (ISN). Careful design of the algorithm for generating these initial sequence numbers ensures that an immediate overlap in sequence number space, between different incarnations of a connection, is prevented. There are other properties required of the algorithm used for initial sequence number generation. To prevent packet manipulation, for a given connection, the ISNs used must be hard to guess for those not involved in the connection [10].

- TCP Time Stamp (TS): the timestamp allows a host to accurately measure the round trip time of a path, and also mitigates problems associated with sequence number wrap-around in links with large bit rate-delay products [8].

The timestamp option consists of two 32 bit fields, TS Value and TS Echo Reply. The TS Value field is set based on the timestamp clock of the sender, and it is into this field that hidden data can be embedded. The only constraints on the timestamp clock are that its tick frequency be between 1 Hz and 1 kHz, and that it be strictly monotonic [6].

The Proposed System

The proposed system is composed of transmitter and receiver hosts that are communicating through the Internet employing TCP/IP protocol suite.

The idea for this work is to manipulate the real traffic by hiding secret data in TCP/IP packet header fields while the traffic still looks normal and no other host can observe the changes that have been made to the packets headers, as shown in block diagram of figure (2).

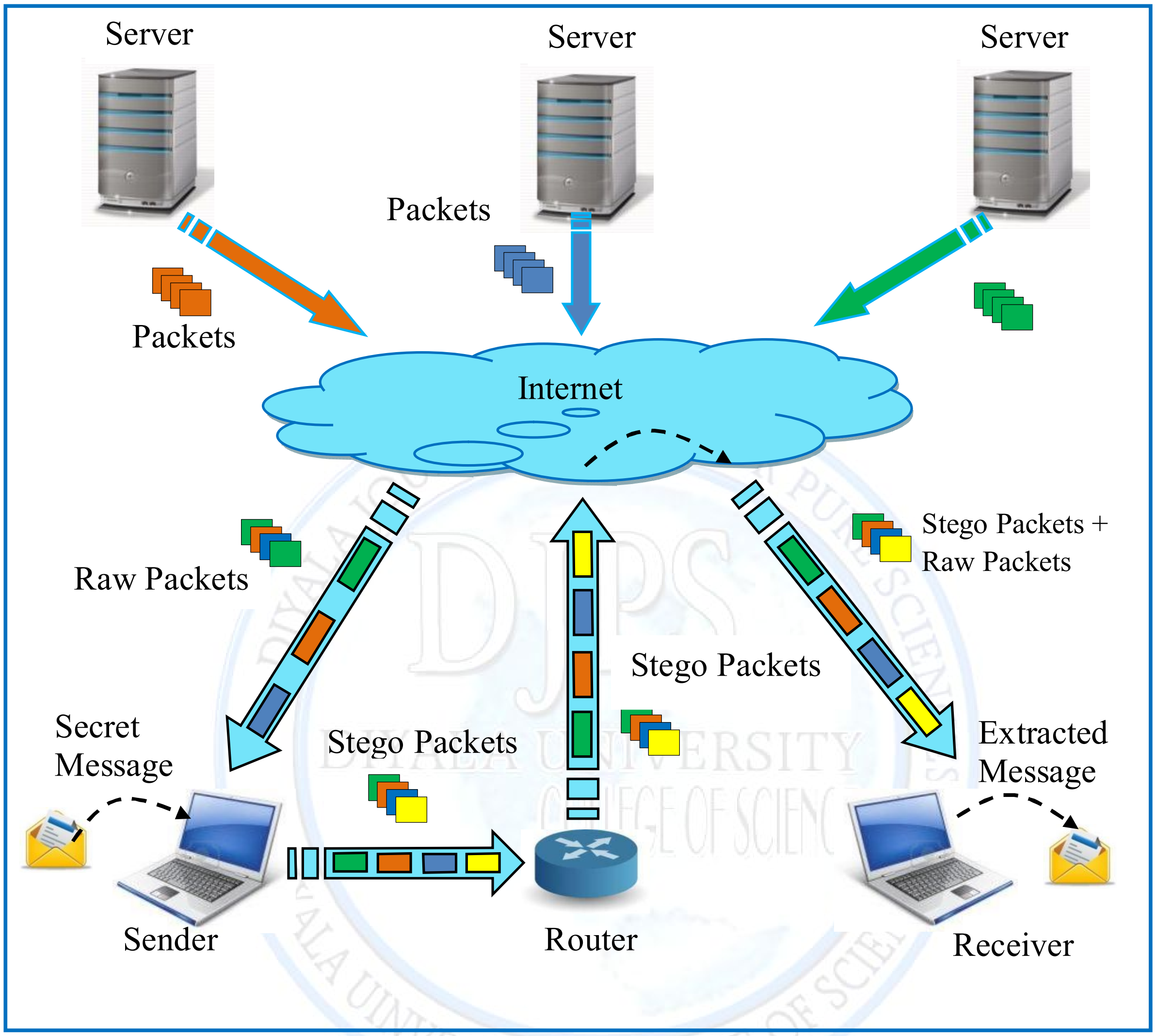

Figure (2) Block diagram of the proposed system

At sender side, packets are captured through the Internet using Wireshark application. Captured packets are stored in offline file in order to be interpreted and analyzed.

Secret characters could be hidden in the header of TCP/IP packets using embedding algorithm. One character is embedded per one chosen packet until the entire secret file is embedded. Then, Stego packets are injected into the network.

At receiver side, packets are received, interpreted and analyzed. Secret characters are extracted using extraction algorithm. Both sender and receiver are communicating through network adapter (Wlan0) with assigned IP address (192.168.1.101) for sender and IP address (192.168.1.100) for receiver.

Different incoming traffics may be received by receiver host from different senders. Therefore, it is difficult for receiver to retrieve the hidden secret message. However, the proposed system is designed to overcome such difficulty and easily extracts the exact secret message, as described in processes of figure (3).

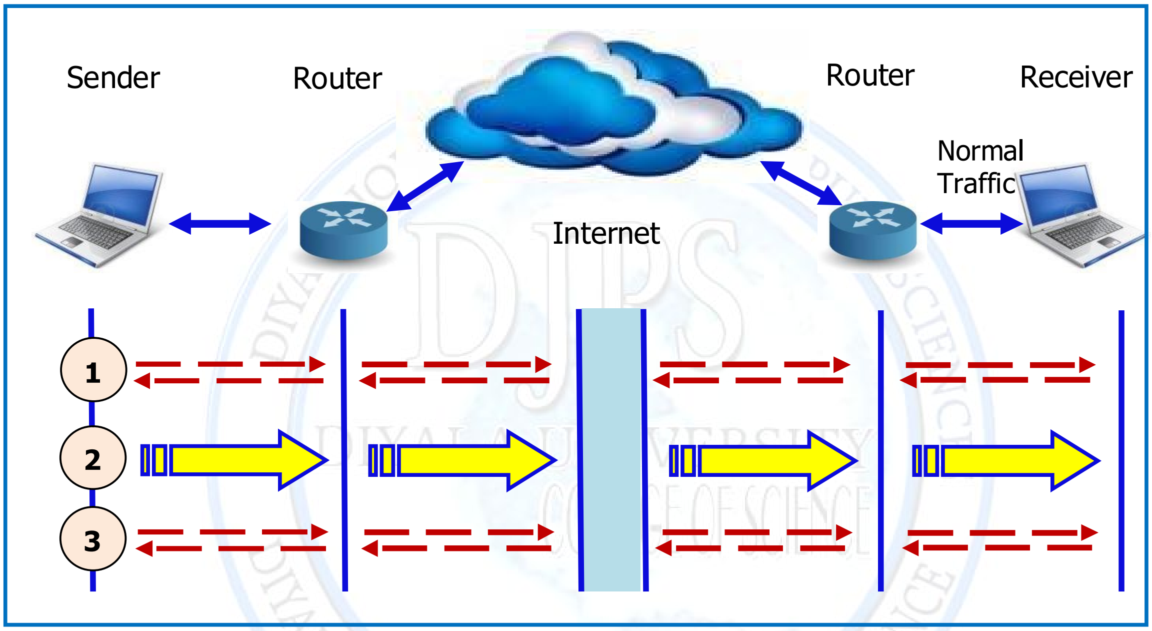

Figure (3) Description of Processes of the Proposed System

The sender [in process 1 of figure (3)] transmits ICMP packets (three pings) to inform the receiver about the next incoming secret message. The receiver is immediately cutout any communication, waiting and only listen to sender’s message in order to avoid traffic mixture, then the sender [in process 2 of figure (3)] starts packets injection where secret message is hidden, while the injected packets are received by the listening receiver. At last, the sender [in process 3 of figure (3)] transmits ICMP packets (three pings) to inform the receiver about the end secret message.

2.1 The Proposed Algorithms

The main function of steganography system is to hide the secret data in an undetectable manner. This could be accomplished by embedding the data within packet header field, IP identification specifically.

The **embedding algorithm** is designed to embed one secret character per chosen packet. The identification field is (16-bit) and the designed embedding algorithm uses the least significant (8-bit) only to make relatively small change to the value of original identification field as shown in Algorithm (1).

Algorithm (1) Embedding Algorithm

```
Input: Chosen packets and Secret Data
Output: Stego packets
------------------------------------------------------------
Step 1: Read character 'char' of embedded file (secret.txt ).
Step 2: Convert character 'char' to its ASCII equivalent value (ch).
Step 3: Read value of 16-bit raw packet identification field (Raw IP ID).
Step 4: Zeroing least significant 8-bit (Temp = Raw IP ID AND 65280).
Step 6: Adding the secret 8-bit ch ASCII  ( New IP ID = Temp + ch )
Step 7: Return packet for injection with  (New IP ID)  field
```

Figure (4) shows a diagram of the designed embedding algorithm. Where x (either 0 or 1) bits represent IP ID of raw packets and C (either 0 or 1) bits represent the secret character to be embedded.

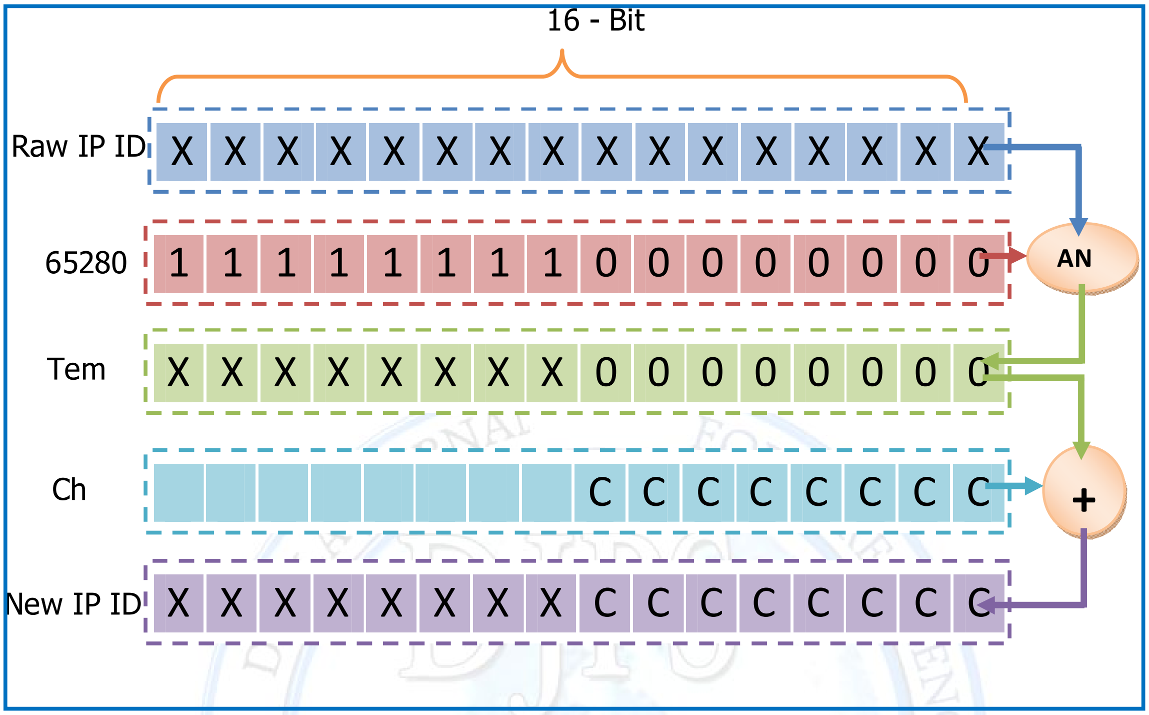
16 - Bit
Raw IP ID
X X X X X X X X X X X X X X X X
65280
1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
AN
Tem
X X X X X X X X 0 0 0 0 0 0 0 0
Ch
C C C C C C C C
+
New IP ID
X X X X X X X X C C C C C C C C

Figure (4) Diagram of the Proposed Embedding Algorithm

The **extracting algorithm** at the receiver side is designed to extract one secret character from each packet. The received packet identification field is a value of 16-bit. The extracting algorithm is shown in Algorithm (2).

Algorithm (2) Extracting Algorithm

```
Input: Stego packets
Output: Secret Data
.................................................................
Step 1: Read value of 16-bit received packet identification field (Rec IP ID).
Step 2: Zeroing most significant 8-bit (Rec char = Rec IP ID AND 255)
Step3: Convert the 8-bit ASCII value (Re char) to equivalent character
Step 4: Save the extracted character to the received file (recsecret.txt)
```

Figure (5) shows a diagram of the extracting algorithm. Where x (either 0 or 1) bits represent the most significant 8-bit of the received packet header identification field bits value, and C (either 0 or 1) bits represent value of the secret character.

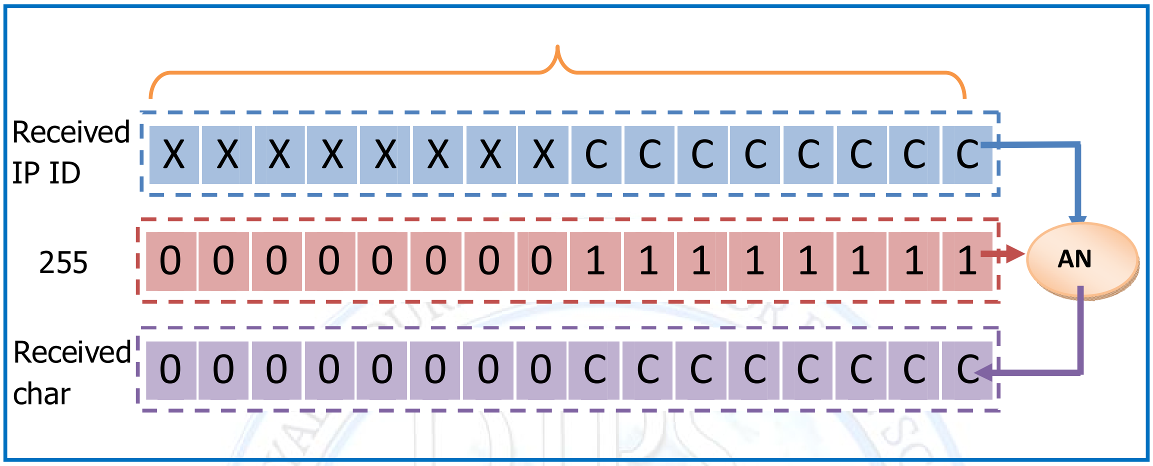

Figure (5) Diagram of the Proposed Extracting Algorithm

Results

Real Internet traffic was captured during this research. The study was conducted to analyze the internet traffic by capturing different type of packets at different situations. The numbers of captured sample packets are (1,000,000 packets). These packets are captured with the aid of (Dumpcap) which is part of the Wireshark application.

Statistics are calculated per session, and average results are presented. Many sessions are captured with different sizes of Pcap file such as (100, 500, 1000, 2000, 3000, 5000 and 10000 raw packets).

The analysis of protocol type field in IP packets prove that most of IP packets are using TCP protocol type as clarified in figure (6) using (100 raw packet).

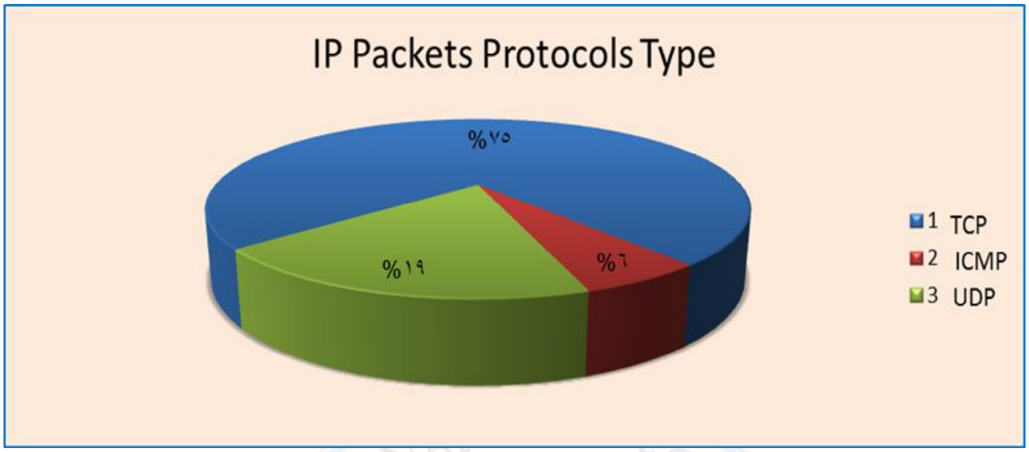

Figure (6) Protocol Types Ratio using (100) Raw Packets

The performance of the proposed system is evaluated and results are presented in this research. Many samples are taken in order to calculate and analyze different experimental results in deferent situations.

The analysis of the captured Internet traffic have been conducted and proven that the minimum length of TCP/IP packet is about 40 bytes (20 byte IP header + 20 byte TCP header + 0 bytes data) and the maximum IP packet length (header and data payload) founded is about 1452 byte as shown in figure (7).

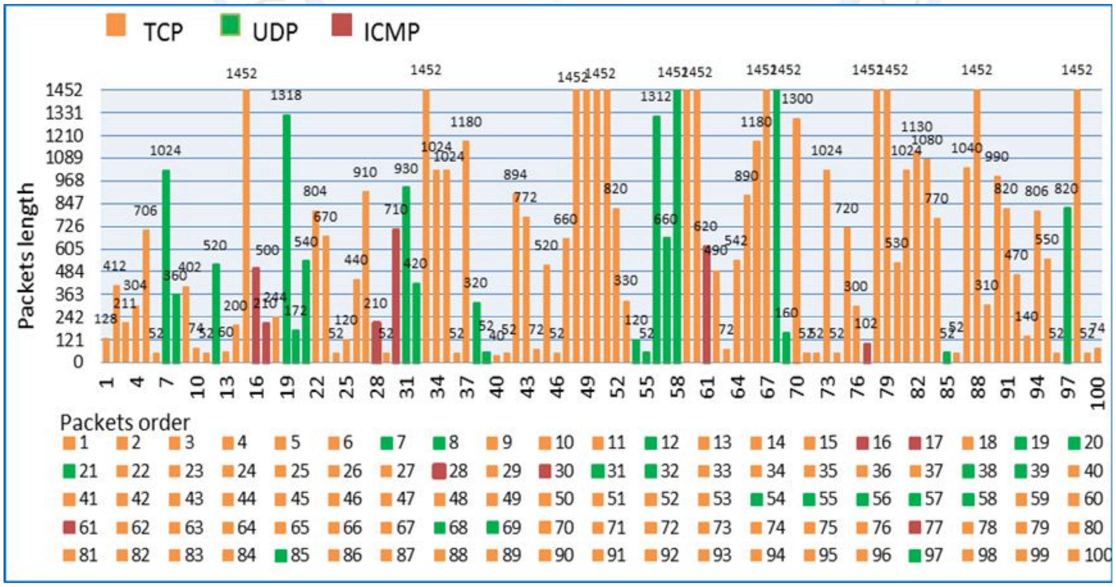

Figure (7) Histogram of Packets Diversity in Sample of (100) Raw Packet

The idea of the proposed system is to embed secret character in IP ID fields of consecutive raw packets and the process continues until the entire secret message is embedded. Therefore, a sample of (100 raw packets) is used as a cover to hide a secret message of (42) characters which is (this is secret file from heart of universe), where each character requires one packet to be embedded within IP ID field. In this case, first ordered raw packets are used in sequence to hide this message, as shown in histogram of figure (8).

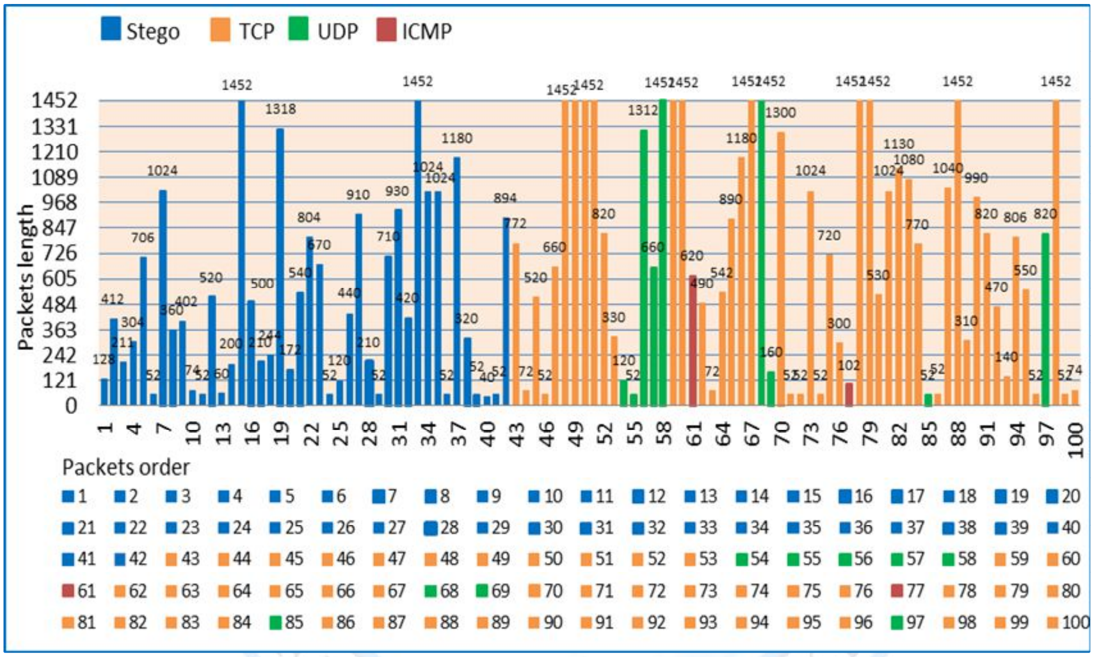

Figure (8) Histogram of Choosing Stego Packets in Sequence

Conclusions

From this work, several conclusions can be drawn as follows:

1. From real traffic analysis, it could be seen that TCP protocol is dominating the Internet that is a rich environment for steganography since there are huge amount of exchanged data.

2. The changes on IP ID are very difficult to detect even when the traffic is analyzed, due to special design of embedding algorithm, which gives transparency to secret data.

3. Network steganography systems usually utilize unused field of protocol header. This work demonstrated that some of the used fields (such as IP ID) in protocol header also could be exploited as cover for steganography data.

4. The packet steganography system requires an acceptable amount of time. The required computation time delay for the proposed system is small and acceptable.

References

1. T. Handel and M. Sandford, "**Hiding Data in the OSI Network Model**", Information Hiding Workshop (IH 1996), Springer Press, vol. (1174) of LNCS, pp. (23–38), Cambridge, UK, May/June, 1996.
2. Stephen Lewis and Steven J. Murdoch, "**Embedding Covert Channels into TCP/IP**", Information Hiding Workshop 2005 proceeding, Cambridge, pp. (1-7), 2005.
3. Craig H. Rowland, "**Covert Channels in the TCP/IP Protocol Suite**", (First Monday Jornal on the Internet), Vol. 2(5), No. (5)., May, 1997.
4. Bender, W., D. Gruhl, and N. Morimoto, "**Techniques for Data Hiding**", IBM Systems Journal , Vol. (35), No. (34), pp. (69-77), 1996.
5. Eric Cole "**Hiding in Plain Sight: Steganography and the Art of Covert Communication**", John Wiley & Sons; Pap/Cdr Edition, 25 April, 2003.
6. Behrouz A. Forouzan, **"TCP/IP Protocol Suite"**, Book from McGraw-Hil Press, 4th Edition, 2010.
7. J. Millen, "**20 Years of Covert Channel Modeling and Analysis**", Proceedings of IEEE Symposium on Security and Privacy, pages 113–114, May 1999.
8. Johnston, P. Kahn D., **"The TCP/IP Protocol Suite**". The Code Breakers, Published by State University Of New York Press, Scribner, 2003.
9. Sebastian Zander, Grenville Armitage, Philip Branch, "**Covert Channels in the IP Time To Live Field**", Australian Telecommunication Networks & Applications Conference (ATNAC), Australia, December 2006.
10. Valeri Korjik, Guillermo Morales-Luna, **"Information Hiding"**, 4th International Workshop, IH Pittsburgh, PA, USA, PP (42-50), April 25–27, 2001 Proceedings.