Steganalysis of Chat based Steganography

Moses Oyaro Okello *

Posted Date: 17 February 2023

doi: 10.20944/preprints202302.0298.v1

Keywords: Steganalysis; Steganography; Timing; Chat

Article

Steganalysis of Chat Based Steganography

Moses Oyaro Okello

Self-Employed, Gulu, Uganda; e-mail: mosesokellomoses@gmail.com

Abstract: Steganalysis is the practice of identifying potential secret communication and taking appropriate action, such as deciphering to uncover the hidden contents or destroying the object containing the hidden information if it cannot be uncovered. At times, it's very necessary to perform Steganalysis due to the fact that steganography is often misused by those with bad intentions, making it a platform for criminal communication. This paper presents a methodology for the detection of timing steganography. The method is based on user behavior during chat, such as the time taken to read, edit, and send text, etc. The method was tested using simulation-based chat software, and it can detect intended timing samples correctly. However, this method is not a very robust one due to some reasons. For instance, it only works well with text-based chat applications, and there are many factors that affect the typing and reading speed of an individual as well as their behavior during chat. These often make the method susceptible to fault detection as it is not easy to set an accurate mean value and deviation to accommodate different typists' normal behavior, which appears as an intended timing.

Keywords: steganalysis; steganography; timing; chat

1. Background

Steganalysis is the practice of discovering hidden secrets within other open, clear information as presented by Nissar, A., and Mir, A. H. (2010) [1], as well as Johnson, N. F., and Jajodia, S. (1998) [2],. Steganography and Steganalysis are both used to improve security because steganography hides information in the presence of an adversary, whereas Steganalysis uncovers or deciphers hidden information in the case of suspected secret communication flows among suspected criminals.

As part of privacy protection act enacted by most Country's constitution, an individual has a right to privacy presented in an article by Kakungulu-Mayambala, R. (2008) [4] and Atuhaire, E. (2021) [5] as well as covert communication article by Okello, M. O. (2022) [3]. However, at times such right are violated by many especially those with bad intention for example criminal might concealed their communication meanwhile plotting to do activities which might be totally against the law or harmful to one another as explained in paper by Elsadig et el. (2022) [6]..

Steganalysis aid in identifying and if possible extract/ recover or decrypt and if it's impossible to decrypt, preferably destroy the suspected payload to disorient such covert communication among parties that would otherwise misuse the right to privacy as presented in an article by Mivule, K., & Turner, C. (2012) [7]. By this, an appropriate measure can be taken early enough.

2. Related Work

This work is basically about detection of timing Steganography in network especially in chat based online application.

For example a work by Gianvecchio etc el. (2007) [8], on network timing detection uses statistical method which detects small variation in signal noise.

A sample of timing steganography work is presented by Liu, G. et el (2012)[9]. Which utilities inter-arrival time of network packets by varying it to hide information.

However, this proposed method is aimed at detecting chat based timing steganography which is presented in a paper by Okello.M (2018) [10], and another paper by Okello.O.M (2021) [11]. These paper uses time interval between two successive time of transmission or texting in the case of online chatting and a single time instance steganography to hides secretes information respectively which

 doi:10.20944/preprints202302.0298.v1

can be used in many platform like online chat application, video time codec, network packets timing, audio timing etc.

In this paper, our main focus is on online chat application timing.

3. Methodology

This proposed method is based on the idea that user behaviors when chatting and attempting to perform timing with an intention to hid information is affected by their intention to send a particular text at a particular instance.

Let time for receiving a message be t_1

Time for checking /reading message be t_2

Time for start typing t_3

Time for stop typing t_4

Time for sending message t_5

Total word count in a text N

Total time for reading text $t_6 = t_3 - t_2$

Total time for typing $t_7 = t_4 - t_3$

Total time from end of typing to sending text $t_8 = t_5 - t_4$

We can set a known average typing speed (mean) μ words per second

With allowable deviation of σ for very fast typist or very slow typist. So $(\mu - \sigma) \leq \mu \leq \mu + \sigma$

Since this methodology reply on typing speed, here we look into some of the few factor such as keyboard arrangements, keyboard types etc. that affects typing speed as explained in the work by William Soukoreff, etc el (1995) [12] which shows that typing speed using Quartz keyboard and other keyboard type affect proficient typist with an average of 30 words per minute meanwhile an inexperienced typist decreases to about 18 words per minute.

3.1. Detection Phase

3.1.1. Typing Speed (t_9)

In this phase, we set an average typing speed of word per second. Therefore.

$$t_9 = N/t_7 \quad (1)$$

So, for a given number of word typed (N), divide by total time spent typing t_6, we get a value which we compare against a set of known average (mean μ) typing speed. But this mean value has minimum and maximum set value to accommodate for slow and very fast typist.

Condition

$$\delta = \begin{cases} t_9 < (\mu - \sigma), \delta^- \\ t_9 > (\mu + \sigma), \delta^+ \\ (\mu - \sigma) \leq t_9 \leq (\mu + \sigma), \delta^0 \end{cases} \quad (2)$$

Interpretation of results:

If Typing result is negative (δ^-), implies that the typing rate is very slow, i.e slower than the average mean value. So it can also mean that the typist is trying to slow down typing speed in order to meet a targeted desired time. Hence possible intend of timing and Steganography detected.

But if Typing result is positive (δ^+), implies that the typing rate is very fast, i.e., faster than the average mean value. So it can also mean that the typist is trying to increase the typing speed in order to meet a targeted desired time. Hence possible intend of timing and Steganography detected. In addition, if typing speed is very high, it could also mean typist just copied and pasted text which was typed somewhere else and just waited for a perfect time to send (timing steganography) and simply copy and past the pre-typed text and send.

And lastly, if Typing result is zero (δ^0), implies that the typing rate is normal, i.e within the average mean value. So it can also mean that the typist is typing at normal speed, hence no steganography detected

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 February 2023 doi:10.20944/preprints202302.0298.v1

However, the problem with this method is that, setting average mean value can be difficult as different typist have different typing rate, hence may lead to fault or inaccurate detection.

3.1.2. Time for Sending ($\boldsymbol{t_8}$)

For this part we target time which a typist finishes typing t_4 to time of sending text t_5

We know that average reading speed of an individual can be set as γ words per minute with a deviation of σ words per minute

i.e.,

$$t_8 = t_5 - t_4 \quad (3)$$

Conditions

$$\delta = \begin{cases} t_8 > (\gamma + \sigma), \delta^- \\ 0 \leq t_8 \leq (\gamma + \sigma), \ \delta^0 \end{cases} \quad (4)$$

Interpretation of Results:

If sending time is negative, δ^-, implies that the typist took some time or delays too much after finishing typing to send the typed text. Hence, possible intend to wait for a specific time to send a given text. And Steganography detected.

But if sending time is zero, δ^0, implies that the sender did not delay that much to send a text after finishing typing, probably was reading to proof read the text before sending or just sent the text without proof reading hence no possible intention, and no steganography detected.

However, the challenges with this method is that sometime, it can be hard to differentiate a typist who unintentionally due to some condition couldn't send a finished text on time. Or the one who would wish to proof read their text after finishing writing.

3.1.3. Reading Speed ($\boldsymbol{t_{10}}$)

This section is only applicable for a received text in situation where there is need to reply for text and also a possible proof reading of typed text after finishing typing before sending, we know that one can spend total time t_6 reading a given text, supposed a known average/ mean for reading text is given as $\gamma = N/t_6$, but according to an article by Brysbaert, M. (2019) [13], an average adult reading speed is about 250 words per minute.

$$t_{10} = N/t_6 \quad (5)$$

Conditions

$$\delta = \begin{cases} t_{10} < (\gamma - \sigma), \delta^- \\ t_{10} > (\gamma + \sigma), \delta^+ \\ (\gamma - \sigma) \leq t_{10} \leq (\gamma + \sigma), \delta^0 \end{cases} \quad (6)$$

Interpretation of results:

If result is negative (δ^-), implies that the reading rate is very slow, i.e slower than the average mean value. So it can also mean that the typist is no reading, just delaying in order to meet a targeted desired time. Hence possible intend of timing and Steganography detected.

But if result is positive (δ^+), implies that the reading rate is very fast, i.e., faster than the average mean value. So it can also mean that the typist didn't read sent text, they simply reply in order to meet a targeted desired time. In this case there could even be possible mismatch of replied text to the received one. Hence possible intend of timing and Steganography detected.

And last if result is zero (δ^0), implies that the reading rate is normal, i.e within the average mean value. So it can also mean that the typist is reading at normal speed, hence no steganography detected

However, the problem with this method is that, setting average mean value can be difficult as different reader implies different treading rate, hence may lead to fault or inaccurate detection.

4. Test Data

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 February 2023 doi:10.20944/preprints202302.0298.v1

In this sub-section, we presents sample data from two different approach. The first one in Table 1 is when typist composes a text to send to someone, and we try to detect this typist typing behavior based on typing abnormalities as a result of their behavior. 1.08333

For testing purposes, we used data from a website blog written by admin [14] which indicates different typing speed based on profession, gender, age etc. The average used here is 50 words per minute or 5/6 words per second in Tables 1 and 2 just for testing purposes. A deviation of about 15 words per minute was used to get the boundary limits

A range of 35 to 65 words per minute about 1.0833~0.5833 words per second.

For reading speed, an average of 250 words per minute with a deviation of about 30 words per minute. About 220 to 270 words per minute and about 4.5~3.6667 words per second. For the highlighted cell in both Tables 1 and 2 indicates detected possible timing.

Table 1. Recorded time activities for Text based Chat and Steganalysis.

t_3	t_4	t_5	N	t_7	t_8	t_9
09:20:02	09:03:12	09:03:14	75	70	2	75/70
10:15:11	10:15:21	10:16:30	63	10	69	63/10
15:00:23	15:12:04	15:14:03	327	701	119	327/701
15:24:34	15:24:46	15:24:49	89	12	3	89/12
16:00:13	16:01:44	16:01:59	65	91	15	65/91

And in the second Table 2, here the target is on someone who received a text message on an online application and replied the text. We also detect the typist behavior to identify any abnormality. Table Cell where it's highlighted gray is the one with some extreme abnormality hence forming a basis of detection.

Table 2. Detection of timing for a Received text and reply.

t_1	t_2	t_3	t_4	t_5	N	t_6	t_7	t_8	t_9
8:30:11	8:50:01	8:52:23	8:52:58	8:53:29	20	142	35	32	20/35
9:10:07	9:10:28	9:10:30	9:11::01	9:11:08	132	02	31	07	132/31
10:02:25	10:20:11	10:21:52	10:21:54	10:21:56	57	41	62	2	57/62
12:10:17	12:15:01	12:15:20	12:15:38	12:15:41	14	19	18	3	14/18
12:31:23	12:31:28	12:32:09	12:35:42	12:36:26	217	41	213	44	217/213

5. Discussion

This method although it prove very good, such as detection of the targets correctly. There are some drawback of the method which still hinder its performance. For example.

At times, user may start writing and paused due to some urgent need to do something else, however this method may not be in position to distinguish this delay and the one caused due to an intended timing, purposely to encode message.

Typing speed varies from one person to another depending on several factors such as someone still learning to type, slow typist, external factors affecting typing speed etc. Hence this may lead to faults detection as it will be difficult to set an accurate mean value and range for normal typing activities. Similarly, reading speed of an individual varies depending on individual familiarity with the language being read, age of reader, complexity of the text such as fiction etc. and other outside factors which affects readers

The method also is also restricted to only online text based chat application and use of timing in different scenario or platform such video based timing and other network based timing may render this method ineffective.

6. Conclusion

This paper presents methodology about Steganalysis which focuses on Chat based timing Steganography.

The method which rely on user/Typist behaviors during typing when chatting such as a recorded time taken to type a given total number of words in a text, time taken after finishing typing to sending text, time taken to read a given text etc.

The method was tested with sample chat on simulated application software and it proved effective as presented, although there are some challenges that hinder the effectiveness of the method which are also discussed in sub-section "Discussion".

I hope as more work on this advances such that a more accurate average mean value for both typing speed and reading speed can be achieved and also the use of artificial intelligence or any other method to detect correlation between two chat text replying to one another will further enhance detection of this method.

At this stage, the method is still not yet hundred percent perfect at detection.

Reference

1. Nissar, A., & Mir, A. H. (2010). Classification of steganalysis techniques: A study. *Digital Signal Processing*, *20*(6), 1758-1770.
2. Johnson, N. F., & Jajodia, S. (1998, September). Steganalysis: The investigation of hidden information. In *1998 IEEE Information Technology Conference, Information Environment for the Future (Cat. No. 98EX228)* (pp. 113-116). IEEE.
3. Okello, M. O. (2022). Optimal Covert Communication Techniques . International Journal of Informatics and Applied Mathematics , 5 (1) , 1-26 . DOI: 10.53508/ijiam.1073205
4. Kakungulu-Mayambala, R. (2008). Phone-tapping & the Right to Privacy: A Comparison of the Right to Privacy in Communication in Uganda & Canada. In *BILETA Conference*.
5. Atuhaire, E. (2021). *Artificial intelligence and the right to privacy in Uganda* (Doctoral dissertation, Makerere University).
6. Elsadig, Muawia & Gafar, Ahmed. (2022). PACKET LENGTH COVERT CHANNEL DETECTION: AN ENSEMBLE MACHINE LEARNING APPROACH. Journal of Theoretical and Applied Information Technology. 100. 7035-7043.
7. Mivule, K., & Turner, C. (2012). Applying Data Privacy Techniques on Published Data in Uganda. In *Proceedings of the International Conference on e-Learning, e-Business, Enterprise Information Systems, and e-Government (EEE)* (p. 1). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).
8. Gianvecchio, S., & Wang, H. (2007, October). Detecting covert timing channels: an entropy-based approach. In *Proceedings of the 14th ACM conference on Computer and communications security*. (pp. 307-316).
9. Liu, G., Zhai, J., & Dai, Y. (2012). Network covert timing channel with distribution matching. *Telecommunication Systems*, *49*(2), 199-205
10. M. Okello, "A New Timing Steganography Algorithm in Real-Time Transmission Devices," *2018 IEEE 18th International Conference on Communication Technology (ICCT)*, Chongqing, China, 2018, pp. 880-884, doi: 10.1109/ICCT.2018.8600103.
11. Okello, M. O. (2021). Transmission of Secret Information Based on Time Instances . The Eurasia Proceedings of Science Technology Engineering and Mathematics , 16 , 209-218 . DOI: 10.55549/epstem.1068612
12. William Soukoreff, R., & Scott Mackenzie, I. (1995). Theoretical upper and lower bounds on typing speed using a stylus and a soft keyboard. Behaviour & Information Technology, 14(6), 370-379.
13. Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of memory and language*, *109*, 104047.
14. Online blog typing speed by admin: https://onlinetyping.org/blog/average-typing-speed.php