Clustering Techniques for Steganalysis

Suchina Parihar
Student ID- 10168164

Master Of Engineering in Electrical and
Computer Engineering

suchina.parihar2@ucalgary.ca

ABSTRACT

Steganography is the art of hiding and transmitting data through apparently inane carriers to conceal the existence of data. The level of visibility is decreased using many hiding techniques in Image Modelling like LSB, Manipulation, Masking and filtering. These techniques are performed by different steganographic algorithms like F5, LSB, JSteg [1]. The skill of discovering the hidden information from the image is called Steganalysis. If we have to monitor a network for leakage of confidential information, there will be many users transmitting objects which must be scrutinized for payload and each actor will transmit multiple objects. To make matters even more complicated, an actor who is guilty of performing steganography will probably behave innocently some of the time, mixing their stegno objects with genuine covers. This is known as Batch Steganography and detecting it is known as pooled steganalysis, a problem presented in 2006 which has still not been addressed successfully [2]. There are so many papers on various data mining and clustering techniques for steganalysis.

Keywords:

Steganography**,** Steganalysis, Clustering, Hierarchical clustering, Agglomerative clustering, Dendrogram, MMD, MDS.

1. INTRODUCTION

1.1 Steganography:

Steganography is a Greek word which means concealed writing. The word “steganos” means covered and “graphical” means writing [15]. Thus, all in all steganography is not only the art of hiding data but also hiding the fact of transmission of secret data. The main aim of Steganography is to hide the secret data in another file in such a way that only the recipient knows the existence of message. In ancient times, the data was protected by hiding it on the back of wax, writing tables, stomach of rabbits or on the scalp of the slaves. But nowadays with the advent of technology, most of the people transmit the data in the form of text, images, video, and audio over the medium. In order to safely transmit the confidential data, the multimedia object like audio, video, images can be used as a cover medium to hide the data [14][15].

1.2 Types of Steganography:

1. **Text Steganography**: The art of hiding information inside the text files is known as Text Steganography. In this method, the secret data is hidden behind every nth letter of every words of text message. Various methods are available for hiding data in text file [16].

2. **Image Steganography**: In image steganography the data is hidden in the images which acts as the cover objects. Pixel intensities are used to hide the data in this steganography type. In digital steganography, images are widely used cover medium because there are number of bits presents in digital representation of an image [17].

3. **Audio Steganography**: Hiding data in audio files is called Audio Steganography. This method hides the data in WAV, AU and MP3 sound files. There are different methods of audio steganography [14]. These methods are

i) Low Bit Encoding ii) Phase Coding iii) Spread Spectrum.

4. **Video Steganography**: It is a technique of hiding any kind of files or data into digital video format. In this case video (combination of pictures) is used as carrier for hiding the data. Generally discrete cosine transform (DCT) alter the values) which is used to hide the data in each of the images in the video, which a human eye cannot notice [14].

5. **Network or Protocol Steganography**: It involves hiding the information by taking the network protocol such as TCP, UDP, ICMP, IP as cover object [18].

2. STEGANALYSIS

The skill of observing the invisible embedded messages in images, audio, video, text as multimedia and protocols is called Steganalysis. The efficiency of any steganalysis method should depend on determining the existence of implanted messages and stego digital images. There is massive quantity of stego hosts which makes this problem really challenging. In this paper, it is proposed to study clustering approaches on steganalysis of images. The main aim of this survey is to present the efficiency of using data mining techniques in steganalysis in comparison to the model based steganalysis approaches [1]. In steganalysis, the problem of detecting hidden data is usually restricted in two ways: only a single actor is considered, and they send only a single object. But such a scenario is unrealistic, in practice, every steganalyst will have to consider multiple actors and each actor will transmit multiple objects. Another limitation is that most traditional steganalysis involves a classification algorithm, which has to be trained on large sets of innocent covers and stego objects [3][4][5]. A new paradigm for steganalysis is proposed, which uses traditional steganalysis features but clustering is introduced rather than classification algorithms. Furthermore, the actors are clustered, based on their aggregate behaviour but not their individual transmitted objects. In this way we remove the need for training, and effectively judge the behaviour of actors by assuming that most of them are innocent. After performing agglomerative hierarchical clustering, the guilty actor(s) should be clustered separately from the innocent ones. We expect this to be less sensitive than steganalysis based on specifically trained classifiers, but more robust and hence more practically applicable. It also attacks the pooled steganalysis problem and is universal, in the sense that an unknown or new embedding algorithm may also be detected, as long as the underlying steganalysis features are sensitive to it [19]. The principle of using clustering to identify guilty behaviour is not new. Clustering has also been used in watermarking to obtain information on the embedding key when a flawed watermarking algorithm creates clusters. However, this application to steganalysis is entirely new [2].

2.1 Classification of Steganalysis:

The steganalysis algorithm may or may not depend on the steganographic algorithm. Generally, steganalysis is classified as follows:

1. Specific / Target steganalysis.
2. Generic / Blind / Universal steganalysis

1. Specific steganalysis: This type of steganalysis is based on analyzing the statistical properties of an image that change after embedding. The advantage of using specific steganalysis is that the results are very accurate. The disadvantage of using this method is that it is cornered to particular embedding algorithm as well as the image format [20][21].
2. **Blind / Universal steganalysis**: In universal steganalysis, the Steganographic Algorithm(SA) is still in the dark. Hence, anyone can design a detector to detect the presence of the secret message that will not depend on SA. Comparing with specific steganalysis, universal is common and inefficient. Still universal steganalysis is widely used than specific ones, because it is independent of the SA [22][23][24].This research focuses on universal steganalysis. It includes the following 2 phases:

a. Feature Extraction.
b. Classification.

a. Feature Extraction: It is a process of creating a set of distinct statistical attributes of an image. These attributes are known as feature. Feature Extraction is nothing but a dimensionality reduction. The extracted features must be sensitive to the embedding artifacts. Some of the feature extraction methods are image quality metrics, wavelet decompositions, moment of image statistic histograms, Markov empirical transition matrix, moment of image statistic from spatial and frequency domain, co-occurrence matrix [25][27][28][29].
b. Classification: It is a way of categorizing the images into classes depending on their feature values. Supervised learning is one of the primary classifications in steganalysis. Supervised learning allows learning under some supervision. In this learning, a set of training inputs that include input features is given as input to train the classifier. After the training, class label is predicted based on the features that are given [26][30]. Steganalysis uses the following classifiers:
1. Multivariate regression.
2. Fisher linear discriminant (FLD).
3. Support vector machine (SVM).
4. Artificial neural network (ANN).

3. OBJECTIVES

The objective of the Project is to solve the problem of data security by actually detecting the theft of data or leakage of any confidential information in a better and efficient manner with the help of data mining techniques specifically Clustering techniques. The other objectives are:

- To study the clustering techniques for steganalysis like Agglomerative Clustering and Maximum Mean Discrepancy Distance.
- To check the accuracy of the proposed method by using multiple-actor, multiple-object setting which clusters the actors.

4. METHODOLOGY

4.1 Agglomerative Clustering:

Agglomerative clustering is one type of hierarchical clustering. Initially, all objects are placed into singleton clusters, the nearest two clusters are combined, and this is repeated until all clusters have been combined and a complete binary tree has been constructed. All that is needed is a method to compute a distance between two clusters, for which there are a number of options [31].

4.1.1 *Agglomerative Process*:

1. Start with a collection C of n singleton clusters
2. Each cluster contains one data point: ci={xi }
3. Repeat until only one cluster is left:
4. Find a pair of clusters that is closest: min D(ci , cj)
5. Merge the clusters ci and cj into a new cluster c i+j
6. Remove ci , cj from the collection C and add c i+j
7. Produces a DENDROGRAM

8. Need to define a distance metrix over clusters

9. Create, traverse distance matrix

3.1.2 *Example:*

Suppose there are 6 singleton clusters P,Q,R,S,T and U. First of all the distance from every singleton cluster to every other singleton cluster is measured. The clusters with the minimum singleton are combined together into a new cluster. This process is repeated until only one larger cluster is left.

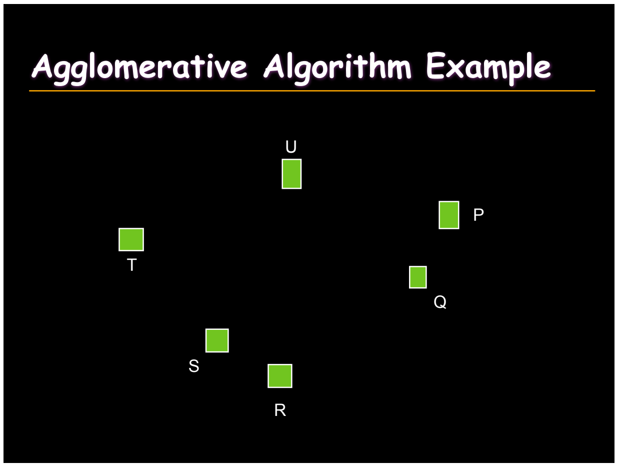

Agglomerative Algorithm Example

Figure 1. Singleton clusters

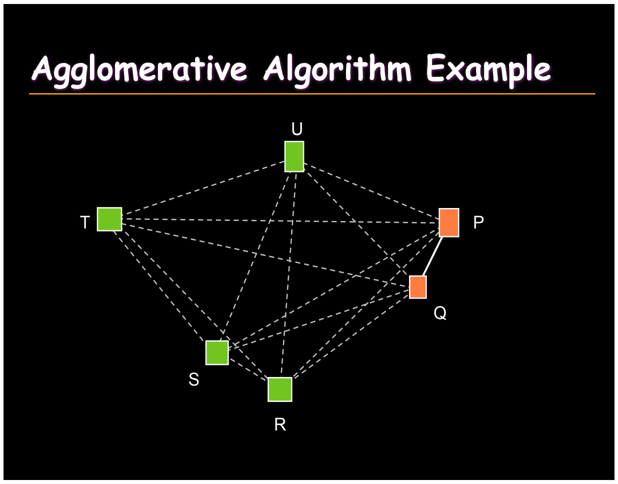

Agglomerative Algorithm Example

Figure 2. Minimum distance between P and Q

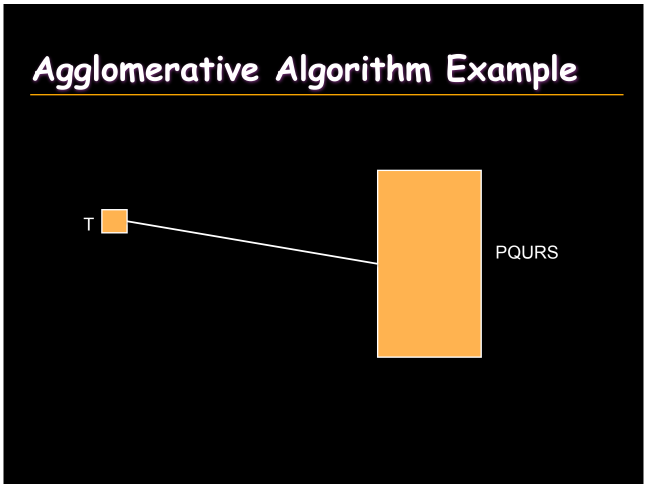

Agglomerative Algorithm Example

Figure 3. Final step of merging the clusters

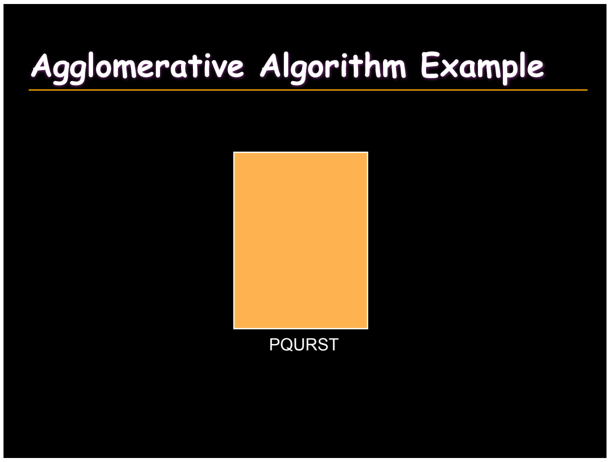

Agglomerative Algorithm Example

Figure 4. One large cluster after merging

3.2 Dendrogram:

It is a graphical device for displaying Clustering Results. In Dendrogram, the vertical lines represent clusters that are joined together. The position of the line on the scale indicates distances at which clusters were joined. The distances between cluster centers indicate how separated the individual pairs of clusters are. Clusters that are widely separated are distinct and therefore desirable

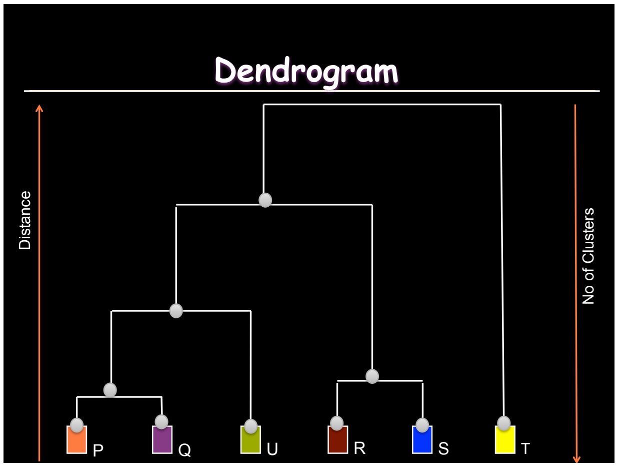


Figure 5. Dendrogram for the example

3.2.1 *Flat Clustering:*

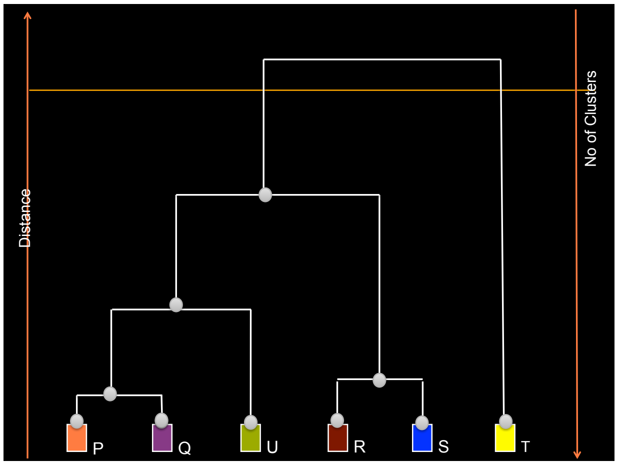


Figure 6. Flat Clustering

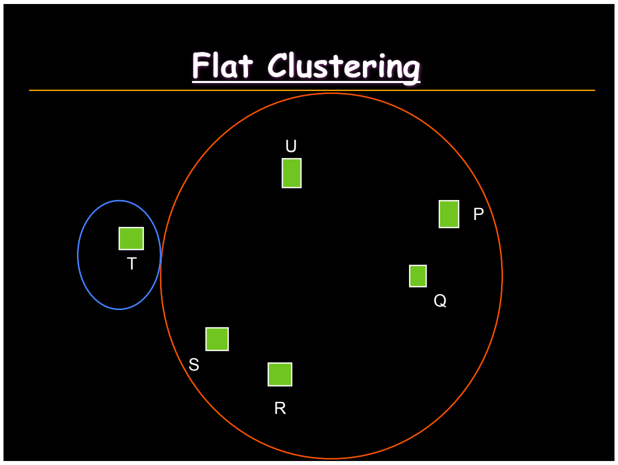
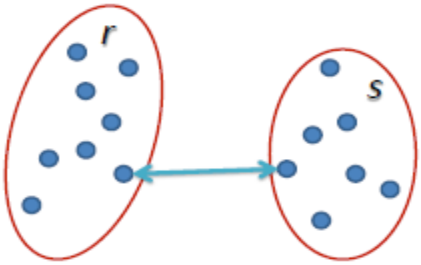


Figure 7. Flat Clustering for the example

3.3 Clustering Distances:

- **MMD measure**: To apply agglomerative clustering to actors in multiple-actor, multiple-object steganalysis, we need a measure of distance between two actors. Supposing that each object has been reduced to a feature vector (of which more later), we can consider the feature vectors they transmit to arise from a probability distribution characterizing the actor's object source and their use, or not, of steganography. Thus, we are given samples from two probability distributions, and require to estimate some sort of distance between the two [12][13]
- **SINGLE Linkage:** In single linkage hierarchical clustering, the distance between two clusters is defined as the shortest distance between two points in each cluster [11] [32].



$$L(r,s) = \min(D(x_{ri}, x_{sj}))$$

Figure 8. Minimum distance

- **COMPLETE Linkage**: In single linkage hierarchical clustering, the distance between two clusters is defined as the farthest distance between two points in each cluster [11][32].

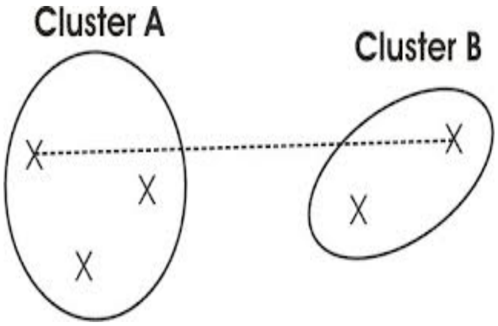


Figure 9. Maximum distance

- **AVERAGE Linkage**: In single linkage hierarchical clustering, the distance between two clusters is defined as the average distance between two points in each cluster [11][32].

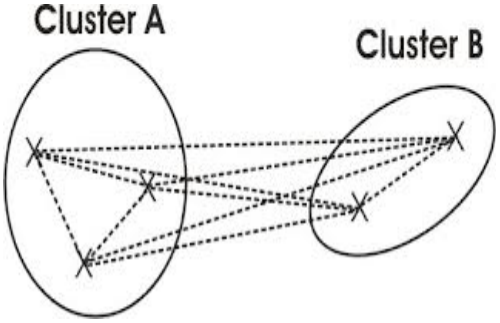


Figure 10. Average distance

- **CENTROID Linkage**: In single linkage hierarchical clustering, the distance between two clusters is defined as the distance of the centroids in each cluster [11][32].

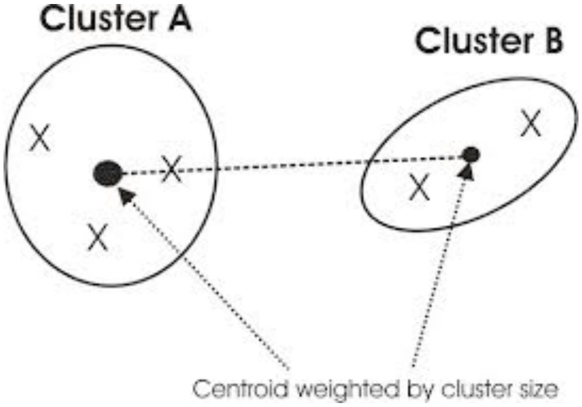


Figure 11. Centroid distance

5. EXPERMENTAL SETUP

This technique is applied to the pooled steganalysis problem. Suppose that there are A actors, each of whom have transmitted N digital images, and we want to identify a single guilty actor from amongst them. In the experiment the actors will be simulated by images taken from different digital cameras, all the images will be JPEG compressed with the same quality factor, and the guilty actor will embed the secret message in some but not necessarily all of the image they transmit. The hierarchical clustering technique is applied with the distance measure between the actors and the expected results are that final agglomeration must be between 1-2 guilty actors and a cluster of innocent actors [8]. Experiments have been conducted on 7 actors sending various types of JPEG images.

Each actor takes N RAW photos and and before transmitting them, converts them to JPEG quality factor 80. For simplicity, constant size images are used by every actor (the default resolution of the camera), but the image size does vary from actor to actor (because the different cameras have different resolutions). To begin with, 50 images are randomly sampled from each actor, and performed no embedding. Then linear MMDs between the image features, for each pair of actors are computed. One way to represent the "shape" of the innocent actors is the multidimensional scaling (MDS) technique, which attempts to locate each actor as a point in a low dimensional space, such that the Euclidean distances between the points are approximately equal to the computed MMD distances. A MDS representation of seven innocent actors appears in the Figure 12. There is no particularly strong outlier. Then we embedded a payload of size 0.7 bits per nonzero DCT coefficient (bpnc) into a randomly chosen 35 (70%) of actor A's images, using the nsF5 embedding algorithm and leaving the other 15 images untouched, and repeated the MMD calculation and clustering. The MDS plot is shown in the Figure 13 and in Figure 15 is the cluster dendrogram, which clearly identifies two clusters consisting of actor A alone, and the innocent users. A has been identified as the guilty actor, despite the detector having information on neither the cover images nor the embedding algorithm. Experiments are repeated with actor B to G as the guilty party, and each was clearly identified [9][10].

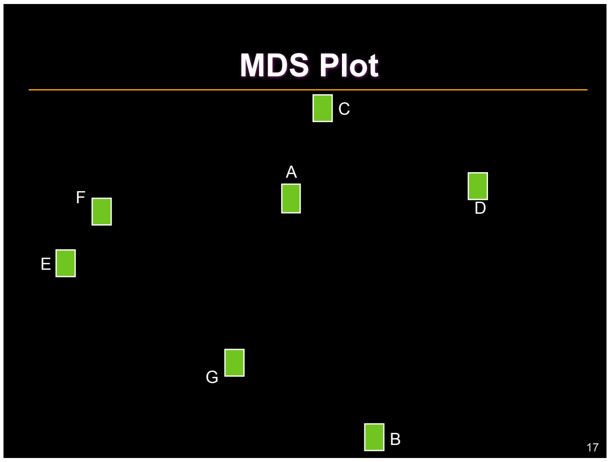


Figure 12. MDS plot of 7 innocent actors

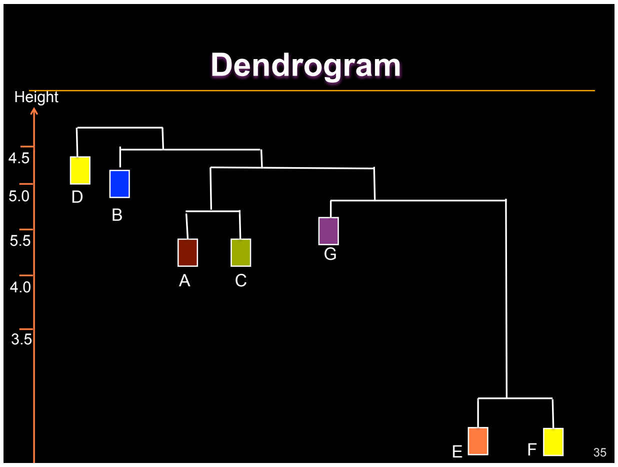


Figure 13. Cluster Dendrogram

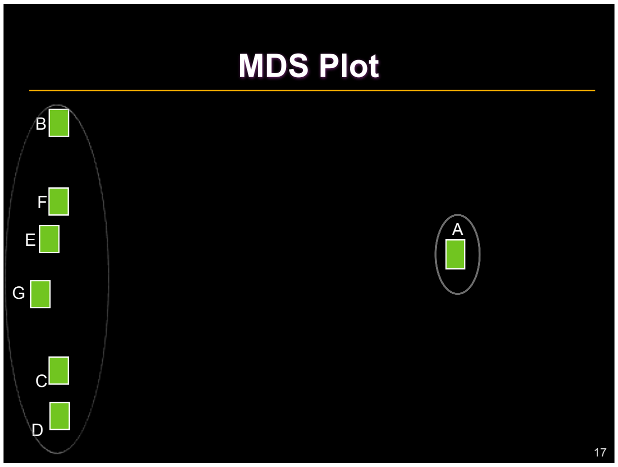


Figure 14. MDS plot when payload 0.7 bpnc is embedded in A

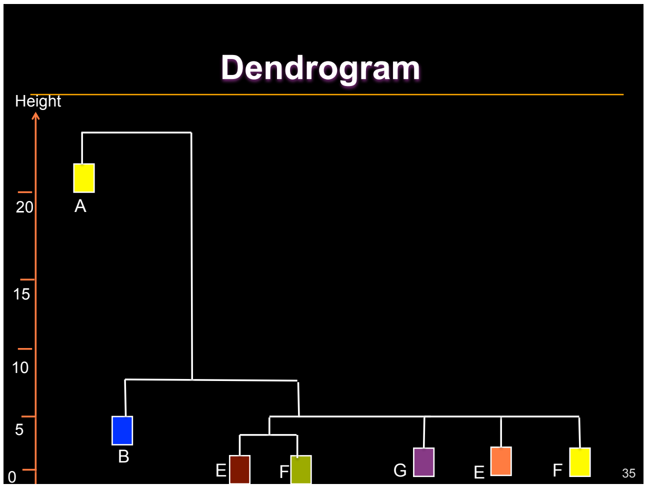


Figure 15. Cluster Dendrogram

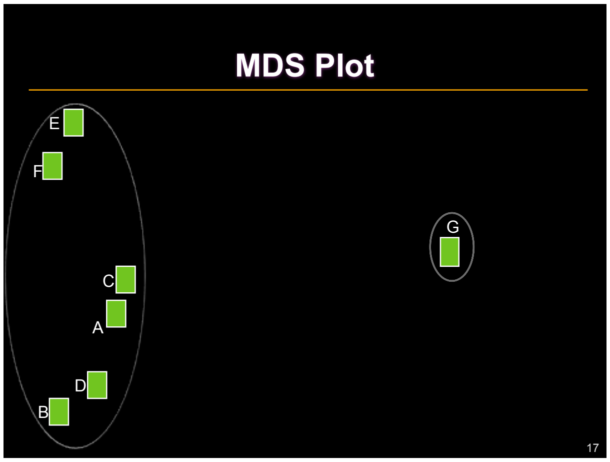


Figure 16. MDS plot when payload of 0.3 bpnc is embedded in G

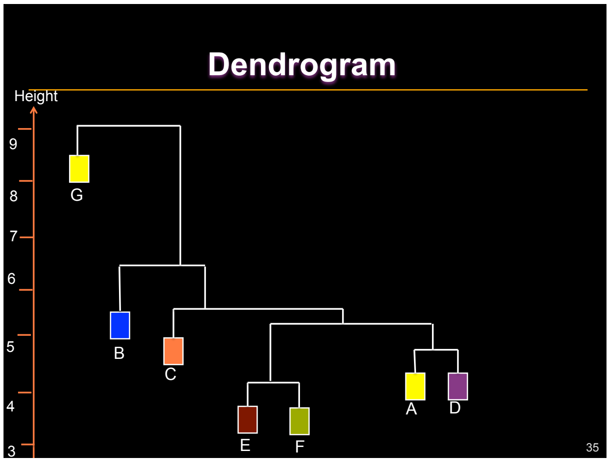


Figure 17. Cluster Dendrogram

To simulate more subtle embedding, the payload is reduced, with the guilty actor using only 60% of images to contain 0.6 bpnc, 50% to contain 0.5 bpnc, and so on down to 10% of images containing 0.1 bpnc. The case of 30% is displayed on the right of Figure 16, with actor G as the guilty party, and still they are identified, although the dendrogram indicates that the distance from G to the rest of the users is not very large compared with the distance between innocent users. These experiments are repeated with different images selected from the database of sources, and different random payloads. The identification of the guilty party is performed by cutting the dendrogram at the final agglomeration: when this consists of a singleton actor merged into a cluster of A – 1, we accuse that actor; when it consists of the merger of a small cluster C with a large complement, we accuse a member of C at random. At payloads of 0.4 bpnc (in 40% of images) or greater, the result is perfect detection [6][8].

6. EXPERIMENTAL RESULTS

Confusion matrices re displayed for the payloads 0.3 bpnc (in 30% of images) and 0.25 bpnc (in 25% of images) in Table. 1, where we observe overall accuracy of 99.9% and 90.3% respectively. Performance falls off sharply for smaller payloads as the guilty actors fade into the innocent cluster and they cannot be identified accurately. These experiments are performed using linear MMD and all seven of the linkage measures from Sect. 2, and with different numbers of images transmitted by each actor: N = 10, 20, 50, 100, 200. This was in order to compare the efficacy of the different combinations. Some of the results are displayed in Table. 2. It was found that the choice of agglomeration algorithm did not make a substantial difference, but that the worst performance arose from tight linkage such as complete linkage, quite good performance from single linkage, and the best performance from centroid linkage. Some of these can be seen in Table. 2. As expected, there is a dependency on N: more evidence allows the detector to make more accurate detection from smaller payloads. When N = 200, linear MMD and centroid linkage allowed perfectly accurate detection of the guilty party with payloads as small as 0.25 bpnc embedded in only 25% of the images: this is an overall relative payload of 0.0625. Even when N is only 20, perfect detection occurs at a relative payload of 0.16. Most steganalyzers require some form of training on both cover and stego images so that the stegalgorithm (or a short list of potential algorithms) should be known, and often their performance is only good if the training set is from a nearly identical source to the images under analysis. In contrast, the clustering method is completely untrained and completely ignorant of the embedding algorithm. This indicate that the rough sizes of payloads detectable by our clustering method are on a par with those detectable by conventional means, at least for the set of seven cameras we tested [6][8].

7. FUTURE WORK

There are two ways to carry this work forward.

First, we could intend to investigate the state-of art in clustering techniques: there are alternatives to agglomerative clustering including hierarchical divisive clustering (an inverse of the agglomerative idea, iteratively breaking the set of actors into smaller clusters) and flat clustering such as k-means. The problem domain can be enlarged to the detection of multiple guilty actors and it would be beneficial to run large-scale experiments. We must also examine the distance metric and Gaussian MMD can also be used along with linear kernels [2][8].

Second, we should consider the features themselves. It is already demonstrated that the feature set is highly sensitive to certain attributes of the cover source, particularly JPEG quantization levels. We deliberately standardized the quality factor of all the images we tested, in the certain knowledge that the method will fail if quality factor varies much because the features will be more affected by difference in source than by embedding. In order to use steganalysis features in the domain of multiple sources, it becomes vital to reduce these variations. Creating feature sets with a more uniform response over different cover sources has not been important in the past because steganalysis was typically benchmarked on single cover sets; clustering can perhaps encourage and inform the design of new feature sets with more uniformity. Indeed most steganalysis literature has, usually implicitly, considered the problem of a single actor sending a single object. We cannot imagine that a steganalyst could be presented with such a simple problem in practice, and an advantage of this approach is that it generalizes the steganalysis problem to a more realistic scenario. Clustering provides a way forwards. Although multiple signals present many opportunities for error (accusing an innocent party, being confused by innocent objects mixed with stego objects) is turned them to our advantage by using the innocent majority to calibrate our expectations, and thus identify the guilty minority. In future work we may be able to quantify the variation in a set of sources, and quantify the embedding signal in terms of payload, and thus begin to consider what steganographic capacity could mean in this context [2][6][7][8].

8. ACKNOWLEDGEMENT

I would like to sincerely thank my course instructor Dr. Reda Alhajj for his constant guidance and supervision.

9. REFERENCES

[1] Nani Koduri,"Steganography: Data hiding using LSB algorithm", February 2011

[2] Ker, A., "Batch steganography and pooled steganalysis," in [Proc. 8th Information Hiding Workshop], Springer LNCS 4437, 265–281 (2006).

[3] Farid, H. and Lyu, S., "Detecting hidden messages using higher-order statistics and support vector ma- chines," in [Proc. 5th Information Hiding Workshop], Springer LNCS 2578, 340–354 (2002).

[4] Harmsen, J. and Pearlman, W., "Higher-order statistical steganalysis of palette images," in [Security and Watermarking of Multimedia Contents V], Proc. SPIE 5020, 131–142 (2003).

[5] Pevny, T. and Fridrich, J., "Multiclass detector of current steganographic methods for JPEG format," IEEE Transactions on Information Forensics and Security 3(4), 635–650 (2008).

[6] Ker, A., "A capacity result for batch steganography," IEEE Signal Processing Letters 14(8), 525–528 (2007).

[7] Ker, A., Pevny, T., Kodovsky, J., and Fridrich, J., "The square root law of steganographic capacity," in [Proc. 10th ACM Workshop on Multimedia and Security], 107–116 (2008).

[8] Pevny, T. and Fridrich, J., "Novelty detection in blind steganalysis," in [Proc. 10th ACM workshop on Multimedia and Security], MM&Sec '08, 167–176, ACM (2008).

[9] Fridrich, J., Pevny, T., and Kodovsky, J., "Statistically undetectable JPEG steganography: Dead ends, challenges, and opportunities," in [Proc. 9th ACM Workshop on Multimedia and Security], 3–14 (2007)

[10] Cox, T. and Cox, M., [Multidimensional Scaling], Chapman and Hall, 2nd ed. (2001).

[11] Beeferman, D. and Berger, A. 2000. Agglomerative clustering of a search engine query log. In Proc. of the Sixth ACM SIGKDD Int'l Conference on Knowledge Discovery and Data Mining, pp. 407–416.

[12] Steinwart, I., "On the influence of the kernel on the consistency of support vector machines," Journal of Machine Learning Research 2, 67–93 (2001).

[13] Gretton, A., Borgwardt, K., Rasch, M., Scolkopf, B., and Smola, A., "A kernel method for the two sample-problem," in [Advances in Neural Information Processing Systems 19], Scholkopf, B., Platt, J., and Hoffman, T., eds., 513–520, MIT Press (2007).

[14] Swati malik, Ajit "Securing Data by Using Cryptography with Steganography" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 5, May 2013.

[15] Moerland, T., "Steganography and Steganalysis", Leiden Institute of Advanced Computing Science, www.liacs.nl/home/ tmoerl/privtech.pdf

[16] F. A. P. Petitcolas, R.J. Anderson, and M. G. Kuhn, "Information hiding- a survey," In Proceedings of IEEE, vol.87, pp. 1062-1078, 1999.

[17] Silman, J., "Steganography and Steganalysis: An Overview", SANS Institute, 2001.

[18] Szczypiorski, K., Steganography in TCP/IP Networks. State of the Art and a Proposal of a New System–HICCUPS, Institute of Telecommunications' seminar, Warsaw University of Technology, Poland,November 2003.

[19] Bhatt, C.A. and Kankanhalli, M.S.(2011). Multimedia data mining: state of the art and challenges, Multimedia Tools and Application, Vol.51, No.1, pp.35-76.

[20] N.F. Johnson, S. Jajodia, Steganalysis of images created using current steganography software, in: Lecture Notes in Computer Science, vol. 1525, Springer-Verlag, Berlin, 1998, pp. 273–289.

[21] J. Fridrich, M. Goljan, Practical steganalysis of digital images-state of the art, in: Proc. SPIE Photonics West, Electronic Imaging (2002), Security and Watermarking of Multimedia Contents, San Jose, CA, vol. 4675, January 2002, pp. 1–13.

[22] Tariq Al Hawi, Mahmoud Al Qutayari, Hassan Barada, Steganalysis attacks on stego images using stego-signatures and statistical image properties, in: TENCON 2004, Region 10 Conference, vol. 2, 2004, pp. 104–107.

[23] M. Niimi, R. Eason, H. Noda, E. Kawaguchi, Intensity histogram steganalysis in BPCS-steganography, in: Proc. SPIE, Security and Watermarking of Multimedia Contents III, vol. 4314, 2001, pp. 555–564.

[24] J. Fridrich, M. Goljan, R. Du, Steganalysis based on JPEG compatibility, in: SPIE Multimedia System and Applications IV, Denver, CO, August 20–24, 2001, pp. 275–280

[25] Wayo Puyati, Somsak Walairacht, Aranya Walairacht. Statistical Moments Based Universal Steganalysis Using JPGE 2-D Array and 2-D Characteristic Function [C]. Proc. of 5th Int'l Workshop on Information Hiding, Berlin: Springer-Verlag, 2006.

[26] Jianjiong Gao, Guorong Xuan, Yunqing Shi est. Steganalysis Based on Moments in the Frequency Domain of Wavelet Sub-band's Histogram for JPEG Images[J]. Computer Applications, 25(10):2170-2172, 2005.

[27] H. Farid and L. Siwei. Detecting Hidden Messages Using Higher-Order Statistics and Support Vector Machines[C]. Pre-proceedings 5th Information Hiding Workshop, Noordwijkerhout, Netherlands, 2002:340–354

[28] J. Fridrich, M.Goljan, D.Hogea. Steganalysis of JPEG Images: Breaking the F5 Algorithm[C]. Proc. of 5th Int'l Workshop on Information Hiding, Berlin: Springer-Verlag, 2002.

[29] Zhexue Ge, Wei Sha. Wavelet Analysis Theory and Application [M]. Beijing: Electronic industry publishing company, 2007.

[30] S. Voloshynoskiy, A. Herrigel, Y. Rytsar, and T. Pun, "StegoWall: Blind statistical detection of hidden data," Proc. SPIE, vol. 4675, pp. 57–68, 2002.

[31] M.S.Yang," A Survey of hierarchical clustering" Mathl. Comput. Modelling Vol. 18, No. 11, pp. 1-16, 1993.

[32] Aggarwal, C.C., Gates, S.C., and Yu, P.S. 1999. On the merits of building categorization systems by supervised clustering. In Proc. of the Fifth ACM SIGKDD Int'l Conference on Knowledge Discovery and Data Mining, pp. 352–356.