Email threading for e-Discovery in Digital forensics

Niladri Sarkar

Digital investigation and forensic computing, University college Dublin.
Email: niladri.sarkar@ucdconnect.ie.

DOI: 10.29322/IJSRP.9.06.2019.p9023
http://dx.doi.org/10.29322/IJSRP.9.06.2019.p9023

Abstract: The process of electronic discovery is very commonly used by digital forensic experts for day today activity to investigate a civil or criminal case. There are several techniques used in electronic discovery keeping the main objective in focus. This paper talks about one of the widely used eDiscovery techniques called Email threading and elaborates the pros and cons of using it during an investigation.

Index terms: Email threading, Digital investigation, Forensic computing, eDiscovery

I. What is e-discovery

Electronic discovery is the process of identifying, collecting and producing information that is electronically stored. This is usually done as a part of an investigation or in response to request for production in a law suit. Electronically stored information many include but not limited to emails, documents, presentations, voicemails, video and audio files. The technologies involved in electronic discovery is often complex because unlike hardcopy evidences the electronically stored data may contain metadata for example time stamps , author etc which has to be maintained along with the original information in order to confirm that the original data was not tampered. Additionally, the volume of electronically produced data is very high which is hard to manage and investigate. [1]

II. Email threading and its advantages

There are several e-discovery analytic technologies and one of the widely used is email threading. It is told that it can cut down review time to more than 60% resulting in both cost and time savings.[2]
An email thread is nothing but a chain consisting of the original message the one that was forwarded and the responses that it received. The tools that help for email threading groups all of the related email messages which in turn makes it easy to review them.

The example given below shows an example of email thread, here gmail acts as a email threading tool which when used in conversation mode and the conversation is between Niladri Sarkar and Nealsrkr.

[1] Complete Discovery Source, 'The Basics: What is e-Discovery'<https://cdslegal.com/knowledge/the-basics-what-is-e-discovery/> accessed on 13th November 2018
[2] LMDglobal,'Email threading saves time in document review',<https://www.ldmglobal.com/2018/08/07/email-threading-saves-time-in-document-review/> accessed on 22 November 2018

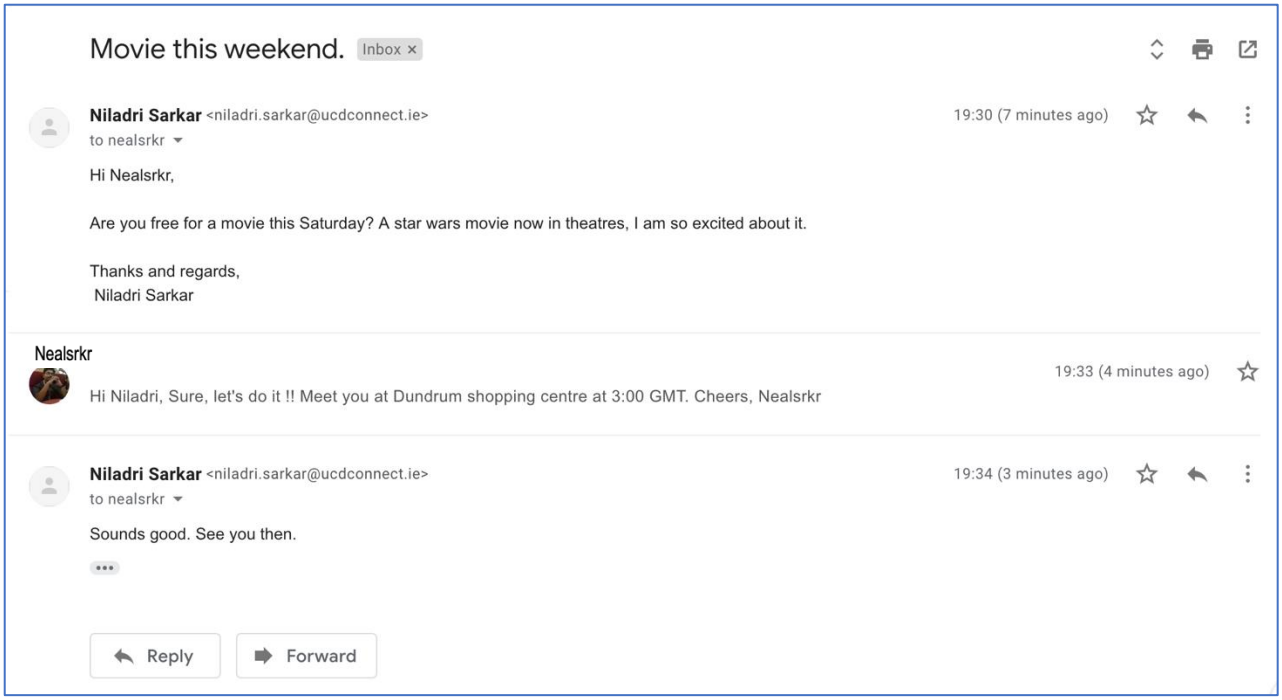
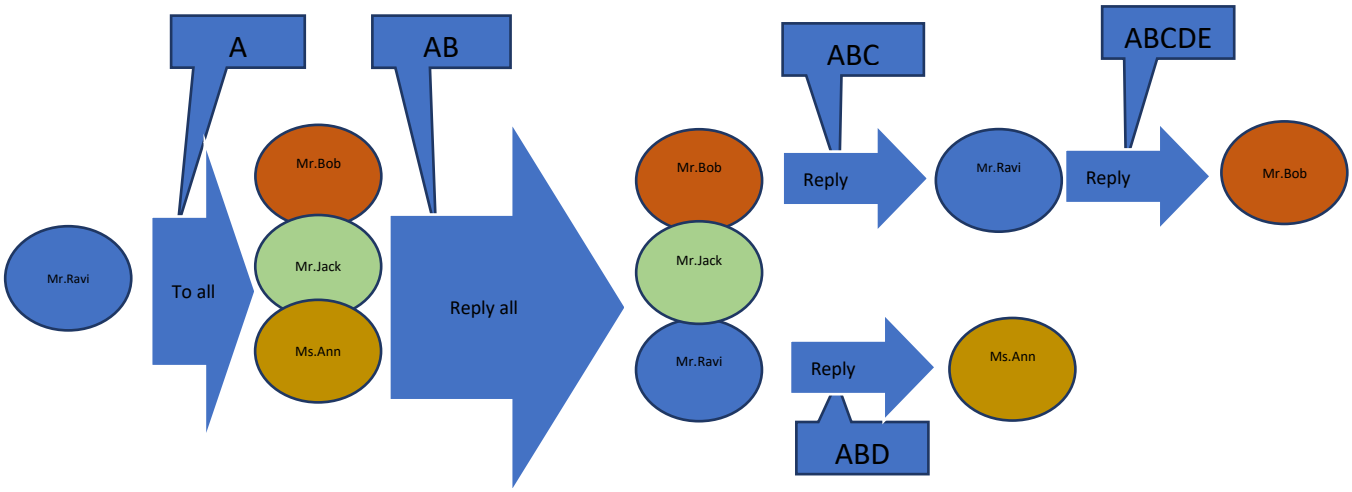


Figure 1: From email box

It is important to have email threading in electronic discovery because it helps to bring in related contents together which would not be otherwise appear together. As an example let us assume that the email invitation for movie from Niladri was forwarded to a third person by Nealsrkr. In this chase without email threading it would have appeared as an absolute new email. However, any email threading tool would group this conversation together and such conversations are usually identified as "near-duplicate" detections. This is because the email that was forwarded by nealsrkr was a copy of the original email sent by Niladri, the only difference is the sender and receiver of the email were different. Sometimes each thread is again sub divided into several other threads following the pattern of flow of messages. The last email is always tagged and reviewed first. The flowchart below is an example of it.

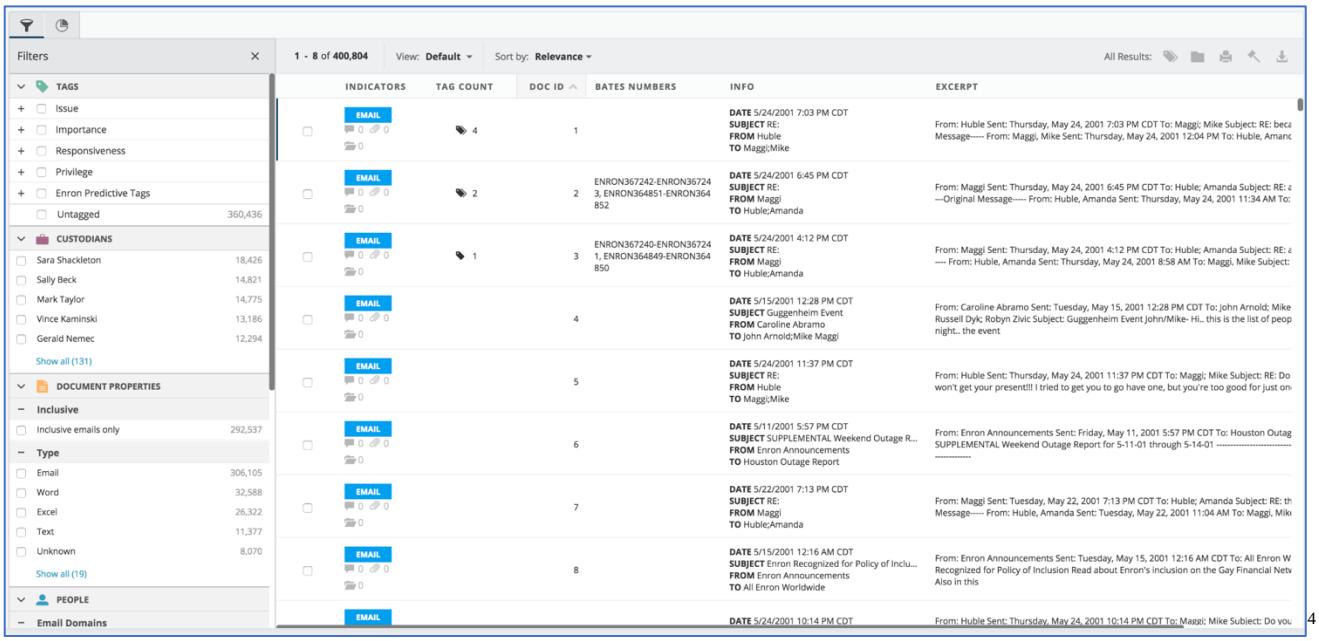


In this example there are:

Type	Count	Description
Thread	1	Initiated by Mr.Ravi
Messages	5	A, AB, ABC, ABD, ABCD
Last Email tags	2	ABD, ABCE

When a document or bulk of email review is down the feature of identifying near duplicates and connecting them to the original data turns out to be extremally useful. In a chunk of information that has been collected the bulk of emails available may not be in correct order as a result if they are not groped well in the first place will lead to insufficiency and create hindrance. This may also led to checking of similar content more than once hence making the process less efficient. Using a email threading tool and as a best practice the last email of an email chain is selected first and all the relevant conversation related to the same email are found in it. Threading an email speeds up the process as an investigator is able to see the content of end to end of all possible data available from a bulk of email along with details like timestamp, sender , receiver and attachments if any. The threading tools are

also equipped with tagging option using which identified piece of evidence can be bookmarked by the investigator and later groped together which can further be used to create an investigation report. A sample view of an email threading tool interface from Disco is shown below.[3]



[4]

When emails are reviewed in the form of a thread it not only helps the investigator to understand the relationship between the sender and receiver but also helps to find out if any conversation or email present in-between are missing, this may help in the investigation process by giving an indication that the interesting or relevant data might me missing or removed and different forensic technique must be initiated to identify or spot the content of the missing piece of information.

III. How email threading is performed

Email threading tools usually categorizes the bulk of provided data first, grouping them in different sections which are the unique data also known as *pivot* by Ringtail tool, the duplicate data and the pre-categorized data which is now duplicate but was original earlier, this is also called as *Duplicate previous pivot* as per Ringtail terminology. An Email thread can have more than one pivot.

Pivot or unique data	The document should meet the following criteria to be a pivot or unique data • It should be the latest content of the thread which has the entire conversation. • It contains the recipients and attachments that are not present in the next document. • The document cannot be thread analysed
duplicate data	• The contents are the exact copy of any other document of the thread at the same level • It contains information which are also available in substituent documents.
Duplicate previous unique	• A new document which becomes a pivot is added to the thread first and the existing pivot is then changes as a duplicate.

Post analysing the entire content the data is classified in the above categories and a thread is formed. Using this analytic approach helps to view a huge volume of data in an organized manner and it makes it easy for the investigator to analyse the same. If there is an email attached to an email it is not considered for email threading, it is just treated like an attachment. An investigator can reduce the volume of documents by removing duplicate ones or my removing any document which he finds to be not relevant. To make the thread analysis more efficient the tools do not take the below prefixes separated by comma under consideration during analysis.

RE, RE:, FW, FW:, FWD, FWD:, Accepted:, Action Requested:, Cancelled:, COMPLETE:, Declined:, NOTICE TO:, Out of Office Autoreply:, Recall:, REMINDER:, Task Accepted:, Task Declined:, Task Request:.

[3] Percipient, 'Email threading? What is it?'<https://percipient.co/email-threading/> accessed on 22 November 2018
[4] Disco, 'Your ediscovery upgrade'<https://www.csdisco.com>, accessed on 22 November 2018

Sometimes it is required to analyse an ongoing email thread, this might result in change in parameter values as they depend on the content of the email.

Below example will help to understand the same:

● Suppose Email 1 is a pivot in a given email thread.

● Email 2 is added on the go and submitted for thread analysis. The new email has the same normalized subject, sender and receiver and also has an extra attachment.

After the thread analysis is completed the results are as follows

● Email 2 will be made as pivot.

● Email 1 is changed to duplicate earlier pivot, as the new document contains all the data it had.[5]

IV. Drawbacks

Although email threading makes the process of eDiscovery easier there are few drawbacks which does not make it cent percent efficient. One of the major drawback is when an email is attached to another email, it is treated as an attachment and its content are not verified like it is done for other emails, this might be necessary in several cases. Also, when the subject is changed and its normalized value is no longer the same the threading tools become inefficient to categorise the same. In the same way if there are multiple email chains running with no subject with not exactly similar contents it is difficult for the threading analyser to categorise such emails. Also, several tools have their own limitation when it is about verifying the attachments, all of them are not efficient to inspect attachments of different file format. It is also a big challenge that all the documents that are suspected to be an evidence has to be downloaded and have to be made available offline to make an email threading tool work on it. This involves requirement of extra infrastructure and time also in this process sometimes there are high chances an calendar event or an email saved as draft can be missed. As mentioned by Microsoft there are chances to loose contents when data is moved to offline platform from one location to another.[6]

V. Application of Email threading in IT Forensics

The main aim of IT forensics is to investigate an incident related to digital technology. It tries to reconstruct the course of events and attempts to identify the individuals involved in an incident. The process may include analysing an email content where email threading plays an important role. Keeping in mind all communications done in an organisation is documented using email also emails are used by authorities to process approvals (this also includes financial departments) email as a document is frequently investigated. One more common incident that is often reported is data theft or passing of data. An employee or former employee may be involved in the act where data has been sent outside the organisation using email and he/she was unauthorised to do so. Considering the huge volume of emails an individual sends or receives it is quite difficult to analyse each of them manually, email threading can here can help to organise the content and present the data in an organised format, this will not only easy the task but also save a lot of time making the process very efficient. Industries are now driven by digital technologies and email is the most impotent mode of communication within an organisation or while communicating with clients, this as a cluster creates a ground for email threading to play an impotent role in IT forensics.

VI. Conclusion

We can conclude by saying that in the field of eDiscovery where identifying collecting and producing information from data which is electronically stored is the major requirement email communication or email as a document is one of the main component of the data to be investigated in most of the cases. The volume of these emails being very high it is difficult to analyse them one by one manually which results in reading the same document more than once. Email threading solves this problem by organising the data and displaying them in an easy understandable visualized format with all meta data like timestamps easy accessible. There are several vendors who provided email threading tolls to make eDiscovery easier for emails. Although each one of them has different terminologies to identify parameters they use almost the same algorithm to create email threads. With advance in technology machine learning plays a vital role to create perfect email threads where the machine trains itself over time to become more efficient. IT forensic investigates on incidents to identify a culprit, frequently is it required to trace emails and reverse engineer to understand what exactly was done which resulted in an incident, email threading is useful in these cases.

[5] Documentation,'Email threading administration',https://www.ringtail.com/documentation/ringtail-9.2/case-admin/threading/thread-admins#MiniTOCBookMark2> accessed on 22 November 2018

[6] Microsoft,'Office 365 meets evolving eDiscovery challenges in a cloud-first world',<https://www.microsoft.com/itshowcase/Article/Content/843/Office-365-meets-evolving-eDiscovery-challenges-in-a-cloudfirst-world> accessed on 22 November 2018

Author: Niladri Sarkar, niladri.sarkar@ucdconnect.ie, MSc. Digital investigation and forensic computing, Department of computer science, University college Dublin.