INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

**ISSN: 2454-132X**
**Impact factor: 4.295**
***(Volume 6, Issue 1)***
Available online at: www.ijariit.com

# Collusion in crowdsourced environments and factors that lead to it: A survey

*Adamu Sulaiman Usman*
*adodass@gmail.com*
*Nigerian Defence Academy, Kaduna, Nigeria*

*Francisca N. Ogwueleka*
*ogwuelekafn@gmail.com*
*Nigerian Defence Academy, Kaduna, Nigeria*

*Abraham Evwiekpaefe*
*contact_abraham@yahoo.com*
*Nigerian Defence Academy, Kaduna, Nigeria*

## ABSTRACT

***Crowdsourcing is increasingly becoming a means by which individuals and organisations seek to solve problems. This is due to the advantages or crowdsourcing such as greater access to experts, lower costs of access to such experts and ubiquitous opportunities. Platforms are, however, plagued with the challenge of collusion between workers. This intern implies that crowdsourcing service providers will be paying out money for no work done leading to losses. This paper presents a survey on crowdsourcing, types of malicious attacks, motivation for workers to collude, factors that lead to collusion and the way tasks are allocated on these platforms.***

***Keywords***— *Crowdsourcing, Collusion, Workers, Task allocation*

## 1. INTRODUCTION

The internet and supporting applications and devices have led to a resultant increase in the amount of communicating devices and indeed and increase in the data generated and created. This data ranges from the text (structured or unstructured) to multimedia (video, images, audio) content on various platforms (sensor networks, machine-to-machine communications, social media sites, cyber-physical systems, and Internet of Things (IoT)) [1]. [1] added that the world produces over 2.5 quintillion bytes of unstructured data every day, with the assertion that over 40 Zettabytes of information will be created and consumed by 2020. With this amount of heterogeneous data being created by various devices, it is clear that we are in the advent of Big Data (BD). According to [2], 97% of companies with a turnover of over $100 million were known to use one form of business analytics or another.

According to [3] big data can be defined by explaining the volume, velocity and variety of data. The term Big data probably started as a casual conversation at Silicon Graphics Inc. in the 1990s [4]. The current hype can be credited to IBM and other leading technology companies in building an analytics market. Big data can be defined as information assets with high volume, velocity and variety. It also has complex heterogeneous data sets that require unconventional techniques and technologies to enable the analysis and use of such data.

The advent of big data and business intelligence has ushered in the use and ability of harnessing the crowd for ideas and providing diverse problem-solving solutions. Crowdsourcing can be seen as a problem and task realization model [5]. The possibility of reaching a large number of participants or researchers over the internet has opened a world of opportunities. Sequel to its success, more researchers are interested in this concept and way of doing things. [5] Stated that the inclusive nature of the Web 2.0 technologies has made it possible for individuals, institutions or organisations to benefit from the use of the crowdsourcing concept. The crowd can be made up of amateurs, professionals, volunteers, organisations etc. Rewards for work carried out can be tangible or intangible.

The model was initially developed in the business environment and has evolved into being used across various fields including medicine, geography and many more. Due to its widespread application, the definition or scope of what is defined as crowdsourcing can be obscure [5].

According to [6], the term "Crowdsourcing" was coined in 2006 by Jeff Howe and Mark Robinson. The official definition was along the lines of crowdsourcing being a company or organisation taking work that was conducted by its staff and outsourcing to an independent undefined network of people. It was later redefined as outsourcing that involved some form of compensation to differentiate crowdsourcing from wikinomics or common based peer production, which involves large groups of people working on the same project such as software projects. [6] added that crowdsourcing describes a way of organising labour in a way where companies can pass out work to be performed by some community largely online. [7] explained that crowdsourcing is an act of taking a task traditionally performed by an individual or employee and outsourcing it by making a public call. This allowed the crowd to provide expertise for what once was the purview of a few professionals. It is the same concept with open source software that is being applied across the business world. Crowdsourcing solves a challenge that has plagued the human species of the previous inability to harness dispersed talent. This system provides a means by which such information is matched to those that require it [7].

According to [8], the quality of crowdsourced data results is influenced by crowd characteristics. This quality has been related to crowd demographics such as gender, profession and age. She added that the failure to produce products that meet the standard of acceptable quality is either as a result of erroneous entry by individuals or a deliberate act to cheat the system. Honest mistakes made by individuals can be handled by careful task design, appropriate task granularity in details of the task process. Identifying and eliminating cheaters from a system requires more stringent quality assurance techniques. She further stressed that quality assurance can be categorized as design-time and run-time. Design-time quality assurance is based on selective worker assignment, good practice of task design and data correction methods. Cost of run-time quality assurance is based on the number of tasks and the likelihood of erroneous submissions [8].

On research conducted by [9] on the identification of non-adversarial collusion in crowdsourcing, it was shown that unintentional results could lead to a significant shift in results obtained. The FINDCOLLUDERS algorithm was developed and proved robust in detecting non-adversarial collusion and showed that, on both real-world and synthetic data, most cases of collusion were detected along with eliminating its side effects. He occasioned that a further look into the incorporation of Machine Learning techniques to detect or investigate collusion was necessary. The focus would be on what types of tasks, or types of raters, based on rater features, were more likely to collude and then use these as standards for a detection model.

According to Chen et al (2018), workers are increasingly colluding to complete the crowdsourced task. The success of crowdsourcing has been witnessed across various tasks such as image labelling and sentiment analysis and also complex tasks such as handwriting recognition, video description, translation and text editing. Chen et al (2018) further explained that the motivation behind collaboration can be two-fold. Some tasks may be quite difficult and could lead workers to collude or workers could be motivated by getting the reward without doing the required work. This could lead to workers providing low-quality answers which could hamper work. Three types of collusion behaviours in crowdsourcing include duplicated submission, group plagiarism and spam accounts. To guaranty that each person will get the same reward, a group of workers could decide to collude and submit the same answer. This is referred to as duplicated submission. Group plagiarism, on the other hand, involves a group of colluders waiting on one person to complete a task and the others duplicate the results. Spam accounts involve workers creating multiple accounts within the same crowdsourcing platform and submitted the same answers multiple times. All three forms of collusion discussed will lead to multiple submissions of the same answers.

Although extensive work has been done on several quality approaches with regards to crowdsourcing methods and results, many issues still remain. For instance, the quality of the outcome of a task depends on several considerations including requesters' requirements, task attributes, crowd interests and incentives, and costs (Allahbakhsh et al, 2013). They further argued that current quality control techniques are largely domain-specific, meaning that techniques that may suit one task may not work very well for another. Finding appropriate approaches to suit various tasks is that challenge that requires more investigation.

## 2. MATERIALS AND METHOD

Semantic searches in Google Scholar were used to retrieve both peer-reviewed and grey literature published on crowdsourcing. This included papers on crowdsourcing, types of malicious attacks by crowd workers, and the motivation for collusion factors that make collusion possible in various sectors and modes of allocating tasks to workers.

This paper reports on collusion in a crowded environment. It also focuses on the definitions of crowdsourcing and malicious attacks, a taxonomy of task allocation and participants and ways in which they participate.

### 2.1 DISCUSSION AND RESULTS

OECD (2011) described collusion in public procurement as a relationship between bidders or participants, who conspire to eliminate the process of competition. The bidders communicate and determine between themselves who will win a contract at a particular time by cover pricing or bid rotation. This influences who wins the bid and when. In many countries, bid-rigging is considered a crime.

With the proliferation of crowdsourcing systems, workers' trustworthiness has become a prominent problem, since it is impossible to tell the large numbers of dishonest workers that may participate in tasks (Yang et al, 2017). For example, some dishonest workers are focussed on receiving the maximum rewards by hastily providing plausible answers, while others intend to boost their trust levels by performing stress-free tasks (Bin, 2015). Due to the existence of untrustworthy answers, requestors often have to ask more crowd workers to answer the same questions to improve the reliability of the answers, which, as a result, greatly increases the economic and time cost for the requestors. This buttresses the importance of the process of selecting trustworthy workers in crowdsourcing systems (Yang et al, 2017). They further explain that crowdsourcing platforms, like Amazon Mechanical Turk 1 and Freelancer 2, have adopted the historical records of tasks to evaluate workers' trustworthiness. Amazon Turk makes use of the overall approval rating to identify trustworthy workers. However, dishonest workers can also easily get high overall approval rates by quickly giving plausible answers or participating in easy tasks. Some researchers are able to carry out trust control mechanisms. For example, Li et al (2014) suggested a general crowd targeting framework that can discover a

group of trustworthy workers based on their characteristics. This technique must also additionally collect the workers' characteristics during the given tasks performed, which makes an additional cost, and can sometimes be impossible to get the complete characteristic information of the crowd workers. In addition, Bin (2015) proposed a context-aware trust model for selecting workers on traditional crowdsourcing platforms, which considers the circumstances like task types and task reward amounts, known as CrowdTrust. Though these methods have considerations such as trust evaluation when selecting trustworthy workers, they entirely neglect the social contexts, like social relationships, social trust and social positions in worker selection.
According to Bin (2015), the results of experiments conducted on datasets provided demonstrated that CrowdTrust can effectively identify dishonest workers. However, the proposed approach may be vulnerable to attack when workers counterfeit fake responses.

In the recent past, crowdsourcing systems have become a significant resource for sharing knowledge, accomplishing tough tasks, collecting opinions, and information gathering. Numerous online systems have made it possible to access a varied range of people to obtain opinions in a short amount of time, regardless of physical location. However, this leaves crowdsourcing for opinions vulnerable to certain threats and challenges such as collusion among participants, either to intentionally skew a rating to a position, or simply to avoid work where a set of colluders' select one person to do the actual work, the others merely copy, and all are paid, drastically reducing much of the benefit to whoever posts and pays for the tasks. An even worse result could be that copying may skew the statistics in crowd ratings, distorting the results for the entity posting the tasks. To avoid getting caught, colluders, instead of exact copying, often make a few changes to their responses such as adding noise, which makes collusion detection more difficult [9]. He added that this type of colluding behavior is seen as non-adversarial, since the goal of the participants is to accumulate payments without doing the actual work, rather than having any malicious intent to favour a certain product over another of competing products or sites.

Other damages caused by collusion exist apart from financial damages; [9] explained that analysis carried out reveals that several statistical metrics such as the mean, median, and variance experience significant shifts as an unfortunate side effect of collusion. Possibilities that could aid in the exchange of information include social networks, crowdsourcing, e-auctions, and e-education. This, in turn, does not negate the importance of these interactions or platforms to solving tasks laid out, if used appropriately. An appropriate use would include one-persona per person; another would be non-collusion among participants. According to [9], previous research on collusion detection was focused on player-player collusion in competitive games. In the case of card games, two or more players can act as colluders.

They could exchange information in order to jointly win the game. E-gaming is games that are supposed to be played under secure e-gaming protocols. This would ensure that some tricks cannot be performed, for example, card exchange, knowledge of opponent cards without their consent, shuffling and dealing the deck for one's own benefit, etc. Therefore, on electronic platforms, if two or more players collude, they can only exchange information. In the real world, it is easier for other players to discover such colluders but, in a virtual world, colluders can use numerous means to collude: mobile phones, social media and other similar platforms. They could even be together while playing (Vallve-Guionnet, 2005). Collusion can also be seen as a covert co-operation between participants of a game. It postures very serious technical, game design, and communal problems to multiplayer games that do not allow the players to share knowledge or resources (Smed, 2007). [9] further states that research has been carried out on cheating by colluding on multiple-choice exams, on plagiarism as a form of collusion, and most recently on on-line reputation and rating systems. [10] explained that the critical aspect of the results obtained from the crowd must be checked for quality. This is required because the background of respondents may be unknown or diverse skills and disparate motivations. The quality of crowdsourced tasks has multiple angles and is contingent on the quality of workers involved, the quality of the task creation process and worker selection. [10] added that detailed studies confirmed that existing crowdsourcing platforms are not robust enough to handle checks and control the quality of crowdsourced data and defend against attacks such as cheating, manipulating task outputs or the extraction of sensitive information from such crowd systems. He explained that crowdsource vendors are not entirely to blame as they may not have the necessary information to approach each task a requester has in mind. An area of concern is the controlling the compliance of rules, regulations and ethical monitoring aspects may ask for bespoke monitoring and assessment techniques in areas such as collusion detection.

## 2.2 TYPES OF MALICIOUS WORKERS

From research carried out by Gadiraju et al (2015), it was discovered that there were different types of malicious attacks on crowdsourced platforms. They started by defining malicious workers as those with ulterior motives to consciously and deliberately go against instructions and expectations of workers. Untrustworthy workers are those unable to provide the correct answer, one or more times, to a given task.

Gadiraju et al (2015) also found that they could classify malicious workers with certain attributes into the following:

- **Ineligible Workers**: Due to the fact that instructions come with tasks to determine proper task completion, those who deviate from these instructions were classified as Ineligible Workers. Pre-requisites may exist before being allowed to answer some tasks and if these are not met, even if data is provided by a worker, it cannot be used ore deemed reliable.
- **Fast Deceivers**: Early on, malicious workers have a tendency to exhibit dubious characteristics by showing an intention to earn fast money via micro-tasks. Some copy and paste the same responses for multiple questions.
- **Rule Breakers:** There was found to be those workers that did not adhere to clear instructions and deliberately did not respond accordingly to each task. Data collected from such workers have little or no value as it cannot be used by the administrator for the intended purpose.
- **Gold-standard preys:** there happen to be some workers that are not malicious that fall prey to Gold-standard questions. This occurs when workers answer such questions wrongly due to exhaustion, boredom or a lack of paying attention.

Gadiraju et al (2015) explained that the use of Gold-standards is the major adopted solution for improved task performance. Gold-standards are simply questions included in the task that the requester already knows the answers to. Therefore, if a worker is unable to provide the right answer to a question or sequence of questions, they are flagged as untrustworthy or unqualified. He discussed that an increase in the use of crowdsourced platforms for solutions will lead to demands for more complex solutions to cheating. This will also result in the need to determine the kinds of malicious attacks that workers can be involved in.

## 2.3 MOTIVATION FOR COLLUSION

Chen et al (2018) explained that the motivation for worker collaboration is two-fold. The difficulty of some tasks cause crowd workers to band together and assist one another. On the other hand, malicious workers want to earn more from crowd platforms while doing less work. A strong angle by malicious workers will result in the submission of low-quality answers, which in turn would reduce the quality of the results those requesters, can make use of. Three types of collusion were summarized with regards to collaborative crowdsourcing.

(a) **Duplicate Submissions:** On the more popular, public crowd platforms like MTurk, rewards are given based on the quality of the results provided by workers. This group of workers may submit the same answers based on the combined intelligence of such a group. This has an adverse effect on the quality of the results submitted by reducing diversity in solutions provided.

(b) **Group Plagiarism:** Some workers aim is to earn as much as possible from crowd tasks, making them more prone to form collaborative groups. When a given worker has completed all the tasks, others just plagiarize and enter the submissions as their own. The worst form of this is workers do not even copy correctly of the initial work is not done correctly as well. They simply forge something that looks like the answer and submits them.

(c) **Spam Accounts:** In this case, some dubious workers register multiple accounts and submit the same answers, multiple times, through these accounts. This type of attack is referred to as the Sybil attack, first discovered or mentioned in distributed systems.

All of these forms of malicious behaviour result in duplicate or multiple submissions made that result in the poor quality of solutions for the requester (Chen et al, 2018). In order to have a good view of collusion, factors that cause and could lead workers to collude were explored.

### 2.4 FACTORS THAT ALLOW COLLUSION

1. **Peers:** Baker (2015) discussed that peers working together can result in improved individual output both in laboratories and in the field. The research was carried out to determine the negative effects of peer settings to determine if this led to cheating or it had a greater motivation for peers to cheat. The investigation carried out also included the consideration of cheating if no additional monetary compensations were made available. The findings showed a significant challenge of cheating in peer settings, though the lack or presence of monetary compensation had little or no effect. It also showed that there is more cheating in peer settings than in individual settings. Scott et al (2011) investigated the influence of peers on one another towards academic cheating. This led to the discovery of occasional to consistent cheaters. It was found that academies with the lowest level of cheating were those with the highest levels of reporting malpractice incidences. There was evidence of large positive peer effects in academic cheating. The model used predicted that one new university cheater is created for every two to three high school cheaters admitted into the university.

2. **Communication:** Is an essential ingredient for collusion to occur. As a result, the elimination of the possibility of communication between parties removes the possibility of collusion (Leonardo, 2011). Luis (2004) made use of a crowdsourcing game to label images. The game was played by thousands of people from different locations. Players were randomly paired so they were unlikely to have any information or knowledge of their partners, which also meant that they would have no way to interact with each other to come up with a strategy for collusion or any form of cheating. This drastically reduced the probability of a partner's ability to cheat.

3. **Nearness to each other:** Luis (2004) stated that several additional steps were taken in order to distance participants from one another. IP addresses of partners must be different from one another in order to make it difficult for them to be paired with themselves. Pre-recorded gameplay was also inserted into the game if collusion was detected in order to scuttle the efforts of colluders. Such collusions were detected by assessing the average time in which players agreed on an image: an extreme reduction in agreement time was indicative of collusion. Also used to deter collusion was the use of taboo words. Once a particular word was used by partners for description, it was not allowed for the duration of their session together.

4. **The complexity of tasks:** According to Tran-Gia et al (2012), from the categorization of crowdsourcing into routine, complex and creative tasks, the detection of cheating in complex and creative tasks is challenging, as the requester must also have the requisite knowledge to understand the results of the work. In both cases for complex and creative tasks, the submissions are also worker specific and this provides a challenge for requesters as it is harder to pin specific outcomes without in-depth understanding of the subject area (Tran-Gia et al, 2012). They argued that current cheat detection mechanisms are either highly specialized or based on gold standard questions (control questions) which are assessed automatically or requires manual checking by the requesters. Manual checking results in a lot of work for requesters and is not wanted. Tran-Gia et al (2012) proposed a method of detecting cheaters in crowdsourcing systems or platforms. They showed that a marginal reduction in cheat detection quality has a great effect on lowering the cost for the cheat detection model where higher prized tasks are concerned. The conclusion was that crowd-based cheat-detection models were cost-effective, reliable and simple to implement. They also have a direct effect on reducing the cost time required for handling cheat-detection manually.

5. **Lack of integrity:** Henry et al (2015) explained that there are many internet services that are dependent on the integrity of users of such systems even when there are reasons for such users to be dishonest. The research conducted evaluated experiments conducted in two different online contexts and evaluated two different methods of cheating and proffered ways of deterring them. Henry et al (2015) argued that there is a lack of evidence about the way to promote honest behaviour in online environments. Methods prescribed for face-to-face deterrence was deemed ineffective on online platforms. An example was provided of a study where participants were asked to adhere to an honour code and this had little if any effect. However, many online exam platforms are still dependent on honour codes to deter cheating. To deter cheating they proposed a technique where there was fair warning about the consequence of cheating and this led to a drastic reduction in those involved in the practice. With research carried out on cheating on Amazon's crowdsourcing platform MTurk, it was discovered that age had no correlation to cheating, however, the number of hours spent on the platform had a direct correlation. The research was also carried out about how task design has an effect on whether or not participants would cheat. It was noted that participants who were more tired, who were in dark rooms, who had less time to think about their actions, who had exercised self-control before the task or those who had felt they had been treated unfairly, were more likely to cheat in increasing measures (Henry et al, 2015).

6. **Vocalization:** According to Van Zant (2014) it has been noted that face-to-face (FTF) interactions provide for more honest responses between individuals as against online or computer-based interactions. In order to test their theory, they made use of a modified version of the deception game theory. They discovered that people are more honest FTF than going through an intermediary. The effect of FTF was found to be greater during the game than before the game. It was also noted that the in-game FTF interactions involved both vocalization and visual access to other participants. As a result, they couldn’t ascertain if either one or both of these factors were necessary to encourage honesty. They explained that there are situations in which vocal communication can be more effective in leading to cooperation among participants than being contingent on visualization alone. This implies that the results obtained could be dependent on whether deceptive information must be vocalized.

7. **Limited time:** Chen et al (2018) postulated that workers are expected to provide independent answers to tasks in order to guarantee answer diversity. To provide more answers with limited time, workers are likely to collude and provide repeated answers, which reduce the quality of accumulated answers. They also stated that there are very few research efforts on the damage caused by collusion. Their researches focused on calculating the effect of worker collusion before and after repeated answers were added. This research was carried out in an effort to determine collusion in result inference. They concluded that as crowdsourcing becomes more popular, there is a great increase in worker collusion, which greatly affects the quality of results obtained by requesters. First, the design of collusion detection was used and then the results were incorporated into various result inference methods.

## 2.5 TASK ALLOCATION

Before any collusion can take place, tasks first have to be allocated to workers. Various authors have demonstrated different conditions and ways in which tasks are allocated.

1. **Push and Pull**: According to Bragg et al (2014), despite there being millions of workers present and numerous amounts of tasks available on crowdsourcing platforms today, the challenge of properly matching the two, known as task routing, remains. They describe the pull and push methods as ways of solving this problem. In the pull mode, such as on Amazon Mechanical Turk (AMT), workers are given the leeway to select tasks based on price, keywords, etc. In contrast, a push-oriented labour market, more common on volunteer crowdsourcing platforms (e.g., Zooniverse), directly allocates appropriate tasks to workers as they arrive. The more information that is available about workers’ history, the better the chances of improved task assignment. For example, a crowdsourcing system might give easy tasks to beginners and route difficult problems to experts. Unfortunately, this potential in itself is a challenge. Existing task directing provided very restrictive assumptions for working crowdsourced platforms to use. They stereotypically assume at least one of the following explanations: (1) tasks can be allocated only sequentially, (2) workers are willing to wait patiently for a task to be assigned, or (3) the quality of a worker’s output can be evaluated straight away. However, an ideal practical task router should be completely unsupervised, since labelling good data is expensive. It must also assign tasks in parallel to all available workers, reducing the waiting period for engaged workers to wait for assignments which could lead to an inefficient platform and frustrated workers (Bragg et al, 2014).

2. **Task Fragmentation**: Haipei et al (2018) postulated a solution for disseminating sensitive tasks to crowds with the possibility of collusion, by breaking a task into various unidentifiable pieces. This makes it harder for anyone or group to decipher the full meaning of a task. These tasks were then disseminated via crowdsourced platforms. Once each sub-task was completed, it was put together by the requester to get a complete picture. The group of workers with the least collusion threshold was selected for these tasks. They inferred that in the future, they would take into cognisance worker quality.

3. **Pay-For-Performance**: In a bid to prevent the participation of insincere workers in crowdsourcing, Shigeo (2014) introduced a pay-for-performance system. Instead of the usually fixed payment system for tasks completed, this method pays workers only based on the quality of the work or results provided. The research included calculating the expected payments for both sincere and insincere workers and then clarified the conditions under which sincere workers were willing to work and conditions which insincere workers were unwilling to work. It was discovered that if they set the fixed amount to zero, it would take a longer for tasks to be completed but not too long. Setting the fixed payment to a small amount was considered more effective than setting it to zero to deter malicious workers (Shigeo, 2014).

4. **Worker Privacy**: Celis et al (2016) worked on a method of protecting the privacy of crowdsourced workers. They argued that there is little or insignificant research in this area. Assigning a confidential job to a single worker may be risky. As such it may be necessary to break up a task into subtasks and given to various workers. This is done with the aim of maximizing the possibility of

privacy. They introduced information loss functions to cater for the amount of information lost during a tasks cycle. The system allows tasks to be requested by workers but also gives the advocates of such tasks the choice of who to send the tasks to.

5. **Worker Relationship:** The rapid development and increase in the use of mobile devices have led to the importance of mobile crowdsourcing as a research focus. Bingxu et al (2019) purported that there are various methods of allocation available. However, few make mention of the combination of social networks and mobile crowdsourcing. Their research focussed on the maximisation of mobile crowdsourcing by considering the characteristics of social networks for crowdsourced systems. The relationship between friends on social networks is applied to crowdsourced systems. A task allocation algorithm is proposed based on the friendship relationship. The GeoHash coding system was used in processing the depth of worker relationship, while also protecting worker privacy. Since spatiotemporal tasks are location-specific, the development of Global Positioning Systems (GPS), and Temperature Probes (TP) has made it possible for mobile crowdsourcing to capture data. Bingxu et al (2019) further explained that a recent study showed that brain activity between friends gets more similar the closer the relationship between them. Therefore, their research focused on the allocation of similar tasks to friends' in order to reduce costs and limit them from having to be in particular locations to participate in location-specific task solutions.

6. **Spatial Allocation:** Hassan (2016) worked on task assignment in spatial crowdsourcing, where tasks arrive dynamically and workers are assigned to these tasks. The objective of the research was to maximize the number of tasks accepted by each worker. The proposed algorithm learns the acceptance rates of each worker and assigns tasks to them accordingly. Dang et al (2013) conducted research and came up with a solution to the spatial allocation of tasks that need to be broken down in subtasks where there is a budgetary constraint. They argued that current system frameworks are inapplicable for such tasks. An alternative was also initiated to minimize the travel costs of participants while maintaining maximum allocation potential.

7. **Visual and Cognitive functions:** Goncalves et al (2017) explained that the allocation of tasks and assignments is important, but often ignored, aspect of crowdsourcing. Their research focussed on the routing of tasks to appropriate workers based on their cognitive abilities. They were able to measure both user visual and fluency cognitive functions and abilities. It was discovered that the performance of crowd workers was in correlation to their cognitive abilities. The proposed model also showed that it was possible to predict the results of crowdsourced activities based on cognitive abilities and proposed the use of such a system for task allocation.

8. **Parallel Routing:** Bragg et al (2014) discussed that an ideal crowdsourcing system would allocate tasks to the appropriate workers. They further stated that the best form of assigning these tasks could be unclear as workers have varied skillsets, tasks vary in complexity and the assignment of multiple workers to a single task could greatly improve results. Their research was focussed on the parallel routing of tasks, by providing iterative methods of task allocation in batches that make optimal use of available workers.

## 3. CONCLUSION

There have been widespread studies carried out on cheating and collusion detection in crowdsourcing. However, only a few if any focus on the deterrence or prevention of collusion, especially with regards to workers' affiliation with each other and factors that promote collusion. To this end, time and other resources are already wasted and lost after tasks have been allocated and carried out, in trying to find out whether collusion took place. Research has also been carried out on the causes of collusion and what facilitates it. Most location-specific research on crowdsourcing has to do with mobile spatial crowdsourcing which is centred on tasks assigned to workers based on their being present at a given location to provide location-specific information. There is also a good amount of research on how tasks should be served, either as a whole or divided up, as a measure to deter collusion and protect the privacy of the task provider. Task allocation was also considered for multiple tasks allocated to one person or multiple persons to the same task. There is however, insignificant research on the way tasks are allocated in order to deter collusion.

This paper has taken a look at crowdsourcing, collusion in crowdsourced environments, factors that cause collusion and various ways tasks are allocated on crowdsourced platforms. The types of malicious attacks range from ineligible workers, fast deceivers, and includes rule breakers and gold standard preys. Fast deceivers are a cause for concern as they form an intentional menace to platforms. The difficulty of tasks and workers with malicious intent form major causes for collusion. The allocation of tasks was an important focus as tasks have to be allocated before any form of collusion, if any, can take place. Factors that lead to collusion include the difficulty of tasks, limited time, complexity of tasks, nearness to one another, peers and the ability for workers to communicate. The surveyed papers highlighted the fact that all these avenues played a role in influencing worker collusion but failed to mention avenues to deter collusion altogether.

## 4. REFERENCES

[1] Sivarajah Uthayasankar, Kamal M, M, Irani Zahir and Weerakkody Vishanth (2016), Critical analysis of Big Data challenges and analytical methods, Journal of Business Research, Vol. 1, No. 7, pp 263–286.

[2] Zafor and Ji. A. S. (2013), Business Analytics: Current State & Challenges, International Conference on Information Resources Management, AIS Electronic Library (AISeL), pp. 2 – 17.

[3] Samiddha Mukherjee and Ravi Shaw (2016), Big Data – Concepts, Applications, Challenges and Future Scope, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, Issue 2, pp. 66.

[4] Amir Gandomi and Murtaza Haider (2015), Beyond the Hype: Big data Concepts, Methods, and Analytics, International Journal of Information Management, Vol 35, No 2, pp 137–144.

[5] Estelles-Arolas Enrique, Fernando González L. Guevara, Raul Navarro-Giner (2015). Crowdsourcing Fundamentals: Definition and Typology. ResearchGate, Vol 3, No 10, pp 1-19.

[6] Paul Whitla (2009), Crowdsourcing and Its Application in Marketing Activities, Contemporary Management Research, Vol. 5, No. 1, pp 15-28.

[7] Howe Jeff, ( 2008), Crowdsourcing: Why the Power of the Crowd is Driving the Future of Business, The International Achievement Institute, Vol. 12, No. 28, pp 1 – 8.

[8] Deniz Iren and Semih Bilgen (2014), Cost of Quality in Crowdsourcing, Human Computation, Vol 1, No 1, pp 1 – 32.

[9] Jansen Peter, Ashiqur R, KhudaBukhsh and Jaime G. Carbonell (2014), Detecting Non-Adversarial Collusion in Crowdsourcing, Proceedings of the Second AAAI Conference on Human Computation and Crowdsourcing, pp 104 – 111.

[10] Daniel Floran, Pavel Kucherbaev, Cinzia Cappiello, Boualem Benatallah and Mohammad Allahbakhsh, (2017), Quality Control in Crowdsourcing: A Survey of Quality Attributes, Assessment Techniques and Assurance Actions, ACM Computing Surveys, Vol 1, No 1, pp 1-45.

[11] Allahbakhsh Mohammad, Boualem Benatallah and Aleksandar Ignjatovic (2013), Quality Control in Crowdsourcing Systems: Issues and Directions, IEEE Internet Computing, Vol 17, No 2, pp 76 – 81.

[12] Bougay Vladimir, Donmez Ayca, Solano Hermosilla Gloria, Subbaraman Balaji, M'barek Robert, Adam Abdoulaye, Bahemuka Stephen, Chinganya Oliver J. M, Eskin Vladimir, … and Santini Fabien (2016), Using Web and Mobile Phone Technologies to collect Food Market Prices in Africa, Garcilaso: JRC Technical Reports, pp 1-88.

[13] Catie Snow Bailard and Steven Livingstone (2014). Crowdsourcing Accountability in a Nigerian Election. Journal of Information Technology & Politics, No. 11, pp. 349–367.

[14] Chen Peng-Peng, Hai-Long Sun, Yi-Li Fang and Jin-Peng Huai (2018), Collusion-Proof Result Inference in Crowdsourcing, Journal of Computer Science, Vol 33, No 2, pp 351–365.

[15] Organisation For Economic Co-Operation And Development (2011). Competition and Procurement. OECD.