Combating Disinformation And Misinformation Through Source Identification And Tracking

Ohwadua, E. O.

Department of Mathematical Sciences, Bingham University, Karu, Nasarawa State, Nigeria
*Corresponding Author:
Email: Emmanuel.ohwadua@binghamuni.edu.ng

Abstract.

In the social media or online networks, the posting and sharing of misleading information or fake news, is done on the fly, but difficult to reverse or control, and worse still, the purveyor of such information is difficult to identify or track; unlike traditional print or electronic media whose authors are well known and can be held responsible for any harm that the information may cause. With disinformation, misinformation and mal-information spreading at the click of a button and at breakneck speed, tech companies are struggling to regulate it on social media, while grappling with public responsibility, definitions of free speech or freedom of expression, and identification of such content and its source. This research piece has provided another prism by which the excesses of the purveyors of disinformation, misinformation and mal-information, otherwise known as fake news, propaganda or misleading information can be curtailed. In the social media platform, the author of content is elusive and may not be known because the content or material does not contain any identification that could help to identify the perpetuator, and even if eventually he/she is finally tracked – probably after the content may have been shared (say) a million times, the harm may have been done. So, the goal of this scheme is to discourage and prevent the creation and posting of harmful material on social media since such content will carry the identification of the author. The identification details of the author of content is acquired at the point of posting the content online on any social media platform, while the integrity of the content is preserved by the use of cryptographic hash function. Fundamentally, this tool, if adopted and implemented by social media platforms, it will shift the focus of the fight against fake news from regulation to prevention and control.

Keywords: *Disinformation, misinformation, mal-information and public-key cryptography.*

I. INTRODUCTION

In today's information landscape, comprehension of the complexity of information dissemination presents an enormous global challenge, particularly given the abundant online information and its impact on the society. This paper builds on the current global research in the areas of information pollution – misinformation, disinformation and mal-information, and proposes a system to ensuring information integrity, source tracking/verification, and prevention/control. Many argue that misinformation, disinformation, mal-information, and hate speech often pollute the information space, threaten peace and security, and disproportionately affect those who are already vulnerable in the society [1]. In the Nobel Prize Summit –Truth, Trust and Hope, hosted by the Nobel Foundation and the National Academy of Sciences, it was stated that the explosion of misinformation and disinformation have weakened public deliberation and undermined confidence in science, even as the world faces interconnecting crises such as war, climate change, and the pandemic as well as other health emergencies [2]. The 21st century has seen the weaponisation of information on an unprecedented scale. Powerful new technology such as deepfake and AI makes the manipulation and fabrication of content simple, and social media and many online platforms dramatically amplify falsehoods peddled by States, populist politicians, and dishonest corporate entities, as they are shared by uncritical publics. The platforms have become fertile ground for computational propaganda, and hate speech, 'trolling' and 'troll armies'; 'sock-puppet' networks', and 'spoofers' – then, there is the arrival of profiteering 'troll farms' around elections [3].

Information pollution, especially online, is affecting the people's capacity to make informed decisions, inciting social divisions and creating mistrust in public institutions, and partly responsible for the consequence of societal crises and the breakdown of public trust in institutions [4]. For instance, recent developments in Europe, such as the war in Ukraine, protests in Kazakhstan, elections in Bosnia,

Herzegovina, and Pakistan, in addition to the current wars in the Middle East are places where misinformation/disinformation is highly prevalent, and we see such phenomenon as one of the key factors impacting negatively in theses crisis. In 2023, during the build-up to Nigeria's general election held in February, ahead of the vote, there were reported explosions of fake news across social media platforms that include divisive content on subjects like religion and ethnicity which were littered across social media networks like Facebook, Twitter, TikTok and WhatsApp [5]. Odanga Madung, a senior researcher on elections at the Mozilla Foundation, told Al Jazeera that the patterns in the information pollution are consistent with observations from Kenya's 2022 election and the trend continues across various democracies in Africa. This misinformation pandemic has increased concerns that could mislead unsuspecting voters, sow political apathy or worse still, lead to violence before, during and after the elections.

This year, 2024, billions of people are expected to go to polls in major elections, totalling around half of the global population, by some estimates – in one of the largest and most consequential democratic exercises in living memory, and the results will affect how the world is run for the next decades [6]. However, false narratives and conspiracy theories have evolved into an increasingly global menace, supercharged by artificial intelligence (AI) which has catapulted misinformation and disinformation efforts, and distorted perceptions of reality to a higher pedestal. With the advent of generative AI, the global 2024 elections have taken on a new dimension, as generative AI systems and other advanced AI technology have made it easier than ever before, to spread misinformation/disinformation through the creation of deepfakes. Most recently in January, voters in New Hampshire, United State, received calls from a voice that sounded like President Joe Biden informing them to stay home and not vote in the primary. According to reports, the AI-generated robocalls might have been created using voice cloning technology from AI voice vendor ElevenLabs [7]. This event, before the New Hampshire primaries, is a reflection of what could come and how much generative AI technology could affect the run-up to the U.S. general election in November and other elections around the globe this year.

1. Information

Information in the digital age have been marked by rapid technological transformations that have completely changed the ways people interact, communicate and access information about our environment [8]. People now possess the entirety of human knowledge in the palm of their hand, and news and information can bounce around the world in seconds. According to Dictionary.com, information is knowledge communicated or received concerning a particular fact or circumstance [9]. Information refers to collected data that has been processed, organized, structured, or presented in a meaningful context, making it useful and relevant to individuals, organizations, society or systems. Information is a critical component of our everyday lives. It is the raw material that drives our communication, decision-making, and enables us to learn and develop. Fundamentally, data and information is the building block of reasoning and knowledge, and without it, we would struggle to make sense of the world around us.

2.1 Sources of information

As was stated earlier, information is a collection of data that has been processed, organized, or structured in a meaningful pattern to convey knowledge, ideas, and/or instructions [10]. It can be communicated through various mediums, such as texts, images, audio, or video, and can be accessed and shared through multiple channels, such as books, websites, and social media platforms. The phenomenal developments in information and communication technology (ICT) over the past two decades have revolutionized the way we produce, consume, and share information – making it easier than ever before to churn out, process and access vast amounts of data remotely from anywhere in the world. The significance of information transcends far beyond our personal lives – it is a critical resource for our societal wellbeing, businesses, governments, and institutions, helping to make informed decisions, and stay safe and competitive in a rapidly changing world. From scientific research to financial analysis, information plays a crucial role in driving development, creativity and innovation across all strata of society.

As we continue to navigate the complexities of the digital millennium, understanding the nature and power of information has never been more intrinsic and valuable. By harnessing its potential and using it wisely, we can unlock new opportunities and solve some of the world's most pressing challenges, on the

other hand, using the information wrongly, can unleash mayhem and trigger upheavals that could result in suspicion, hate and even war. At its core, information is a representation of reality, and it is used to convey knowledge about our environment and the world around us. It can be factual, subjective, or even fictional, and can take many different forms depending on its purpose and audience – either for good or for bad. The quality and accuracy of the information can significantly impact individual and collective outcomes, making it crucial to critically evaluate and verify the sources and reliability of the information we consume and share. As technology continues to evolve in a fast pace in this 21st century, the amount of information available, particularly on the internet is growing exponentially. This has created new challenges around confidentiality, integrity and availability as well as processing, and using information effectively and appropriately.

1.2 Disinformation

Disinformation is a false or misleading content that is deliberately created and spread with an intention to deceive or secure economic or political gain, and which may cause public harm to social group, communities, organisation or country [11]. The spread of disinformation can have a scale of unwholesome consequences, such as threatening our democracies, polarising debates, and putting the health, security and environment at risk. The creation, sharing, and consumption of disinformation and fabricated content on social media and other online platforms is a growing concern, especially with the ease of access to such sources, and the lack of awareness of the existence of such inaccurate or false information. Disinformation can be spread by state or non-state actors and could affect a broad spectrum of human rights, undermining responses to public policies or amplifying tensions in times of emergency or armed conflict [8].

1.3 Misinformation

Misinformation is information that is false, but not created with the intention of causing harm. Misinformation can hold sway for years, even after the facts are set straight and it spreads faster than true information because of its social and emotional qualities. Research shows that misinformation can be "sticky" if it's frequently liked, commented, or shared – or if it evokes feelings of fear or mistrust [12]. Misinformation spreads rapidly across social media and other online platforms, constituting risks to individual health and societal well-being and distorting inclusive news and narratives for a well-informed public. Several researches on the psychology of misinformation have proliferated in recent years, yet many questions remain about how and why misinformation spreads, how it affects behaviour, and how best to counter it – answering these questions well depends in part on how misinformation is defined – it can include inaccurate news, conspiracy theories, disinformation campaigns, propaganda, and slanted reporting [13].

1.4 Mal-information

Mal-information is information that is based on real facts, but manipulated and disseminated to deceive or inflict harm on a person, group, organization or country. It often stems from the truth but is exaggerated or contextually misrepresented in ways that can mislead and cause potential harm [14]. It could be a deliberate publication of private information for personal or corporate rather than public interest, such as revenge porn or conscious change of context, date or time of genuine content. It is often associated with propaganda, hoaxes, and disinformation campaigns.

1.5 Hate Speech

Hate speech refers to any kind of offensive communication in speech, writing or behaviour, that attacks or uses pejorative or discriminatory remarks with reference to a person or a group on the basis of who they are – based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor [15]. Typical hate speech involves epithets and slurs, statements that promote malicious stereotypes, and speech intended to incite hatred or violence against a group and can also include nonverbal depictions and symbols [16]. The growth of hateful content on online platforms have been aided with the rise of easily shareable disinformation enabled by digital tools. This raises unprecedented challenges for our societies as government institutions struggle to enforce national laws in the virtual world's scale and speed. Critics of hate speech argue not only that it causes psychological harm to its victims, and physical harm when it incites violence, but also that it undermines the human rights of the victims. Hate speech therefore poses a challenge for modern liberal societies, which are committed to both freedoms of expression and social justice. The

relative permanence of hateful online content is also problematic, as it can resurface and (re)gain popularity over time.

1.6 Prevalence of Dis-, Mis-, and Mal-information on Social Media

Dis/Mis/Mal-information which I may group together here simply as fake news or misleading information are not new but rather have become increasingly more powerful as they are driven by new technologies and rapid online sharing or dissemination. The sharing or dissemination of content – text, images, videos, or links online, for example, allows information to go viral within hours if not even minutes. The principal difference between misinformation, disinformation and mal-information is the intent of the actor, entity or author creating the content. Fake news, propaganda or misleading information has been described as a global harm, and the amount of misinformation encountered by people on social media or online platform is difficult to determine. According a report, two thirds of EU citizens report coming across fake news at least once a week [17].

Estimates indicate that it accounts for 0.2% to 29% of overall news consumption, but the proportion may be higher for specific groups or for topics such as health [13]. The problem with current estimates is that they tend to be platform-specific, constrained to text-based information (vs. images or videos), based on limited public data, and insensitive to the fact that some groups or communities are disparately targeted. Whilst social media platforms offer a wealth of information, communication possibilities, and entertainment, inaccurate and misleading content is a persistent concern on online networks. The posting and sharing of misleading information or fake news, is very easily done, but difficult to reverse or control, and worse still, the purveyor of such information might be difficult to identify or track, unlike traditional print or electronic media whose authors are well known and can be held responsible for any harm that the information may cause. With fake news spreading at the click of a button and breakneck speed, tech companies are struggling to regulate it on social media, while grappling with public responsibility, definitions of free speech or freedom of expression, and identification of such content and its source [18].

1.7 Efforts at Combating Disinformation and Misinformation on Social Media

In response to the urgent need to combat rising disinformation and misinformation, several efforts to combatting the wave of fake news or propaganda have been gaining traction over the years – some by various national government – mostly through legislation, others are by tech companies such as Facebook, X (Twitter), YouTube, etc., finding ways to detect and remove fake news from circulation. For instance, according to Meta, and I quote, “we are stopping false news from spreading, removing content that violates our policies, and giving people more information so they can decide what to read, trust and share” [19]. In the case of tech companies like Meta, the harm may have been done before the detection and eventual removal of the offensive material. So, the capacity to respond in real time to online fake news is more than a safe-guarding tool, however, it is also an important democratic competence in its own right. Although fake news on social media can seem like a problem without a clear solution, the subject is widely acknowledged by experts in the tech industry as a highly concerning one. The tackling of false and misleading content online is actively receiving attention and action from professionals and policymakers. In Europe for instance, the Digital Services Act, aimed at creating safer online communication and environments, intends to hold large platforms accountable for harmful content, as well as carry out risk assessments, independent audits, and overall transparency regarding the usage of algorithms [18].

Equally, in the United Kingdom, a proposed Online Safety Bill is calling for more consideration towards users from online platforms and aims to target trolling, harmful content, and internet fraud, among other pressing online issues. In the United States, in a white paper released by Cyberspace Solarium Commission (CSC), the federal responses to disinformation campaigns have largely focused on law enforcement action and sanctions against actors attempting to interfere in the media environment, to provide a sustainable and agile approach that can mitigate harm [20]. Also in the report, the CSC recommends teaching of digital literacy and reinvigorating civic education to counter the erosion of trust in democracy and democratic institutions that is so often the goal of disinformation. It is no doubt a general global consensus that media and information literacy (MIL) have a good effect on the ability to identify fake news, disinformation and misinformation, and sharing intentions, but unfortunately, these approaches are yet to

completely combat this menace. The hearts and minds of all kinds of consumers are valuable prizes, which means misinformation sources are continually coming up with new ways to deceive and mislead – that makes it increasingly difficult for the average media consumer to identify potential false news [20].

2. New Approach to Combating Dis-, Mis-, and Mal-information on Social Media

Over the years, there is no doubt that combating fake news has been receiving a lot of attention from various government and tech companies all over the world. Despite these efforts, the prevalence of fake news has been unabated – no thanks to new and sophisticated AI tools that helps in the creation and manipulation of content either text, image, audio or video, and the ubiquitous social media or online platforms that help to spread the content in matter of minutes to subscribers, and within hours, the information has gone viral all over the world.

Our new approach may not necessarily be a one-stop solution to the menace of fake news, but we are convinced that it will add to the multilayer approaches already available in combating false news. The goal of this approach is to help to track, verify and identify the source and possibly the author of fake news on online or social media networks. More like the traditional print and electronic media, the source and identity of the author are usually not in doubt, and hence any actor that is responsible for the production of fake or malicious content can be held responsible for the harm and damage to the victims or society. Our approach is to replicate this non-anonymity of the author of false or malicious information on traditional media to make it equally mandatory for the purveyor of fake news on online and/or social media platforms to similarly remove the veil of anonymity to make sure that the source and identity of the actors can be tracked and verified.

3.1 Design Features

The design features of our scheme include the following:

3.1.1 MAC Address

A Media Access Control (MAC) address is a string of characters that identifies a device on a network. It is a unique identifier assigned to most network adapters or network interface cards (NICs) by the manufacturer for identification and use in the Media Access Control protocol sub-layer. The NIC is essentially a computer circuit card that makes it possible for your computer to connect to a network. The MAC address is a 48-bit binary value expressed as 12 hexadecimal number (6-bit binary number), which is mostly represented by Colon-Hexadecimal notation. The First 6 digits (say 00:40:96) of the MAC Address identify the manufacturer, called the OUI (Organizational Unique Identifier) while the rightmost six digits represent Network Interface Controller, which is assigned by the manufacturer. Within a network, each MAC address ensures accurate delivery of information which means that devices on the network can be easily identified and managed. Manufacturers may identify a MAC address by other names, such as the physical address, hardware ID, wireless ID, and Wi-Fi address.

Characteristics of MAC Address

- MAC addresses are used in LAN (Local Area Network) environments to identify devices and allow communication between them.
- The first 3 bytes of a MAC address represent the manufacturer ID, while the last 3 bytes represent a unique identifier assigned by the manufacturer.
- MAC addresses are burned into the hardware of a network interface card (NIC) and cannot be changed, except in some rare cases where the manufacturer has provided a specific tool to do so.
- MAC addresses are often used in conjunction with ARP (Address Resolution Protocol) to resolve IP addresses to MAC addresses for communication on a LAN.
- Some operating systems, such as Windows and Linux, allow you to view the MAC address of your network adapter through a command prompt or network settings.

3.1.2 IP Address

An Internet Protocol (IP) address is a unique identifying number assigned to every device – computers or nodes connected to the internet or local network. It is a string of numbers separated by periods and are expressed as a set of four numbers – each number in the set can range from 0 to 255. So, the full IP addressing range goes from 0.0.0.0 to 255.255.255.255. IP addresses are not random – they are

mathematically produced and allocated by the Internet Assigned Numbers Authority (IANA), a division of the Internet Corporation for Assigned Names and Numbers (ICANN). Every device that communicates over the internet or via local networks share information to a specific location using the IP addresses. All devices with an internet connection has an IP address, whether it's a computer, laptop, IoT device, or even toys. The IP addresses allow for the efficient transfer of data between two connected devices, allowing machines on different networks to talk to each other. IP addresses have two distinct versions or standards – IPv4 and IPv6. The Internet Protocol version 4 (IPv4) address is the older of the two, which has space for up to 4 billion IP addresses and is assigned to all computers. In other to accommodate more devices to be connected to the internet, the IPv6 was developed known as the Internet Protocol version 6 (IPv6) which has space for trillions of IP addresses, and accounts for the new breed of devices in addition to computers.

There are several types of IP addresses, including public, private, static, and dynamic IP addresses:

Types of IP Addresses

An internet service plan for every individual or business will typically have two types of IP addresses – private IP addresses and public IP address. The terms public and private relate to the network location – a private IP address is used inside a network, while a public IP address is used outside a network.

- **Private IP Address:** A private IP address, or internal-facing IP address, is assigned by an office or home intranet (or local area network) or router to devices. This includes computers, smartphones, and tablets but also any Bluetooth-enabled devices such as speakers, printers, or smart TVs. Nowadays, with the growing internet of things (IoT) devices, the number of private IP addresses have at home is probably growing. The router needs a way to identify these devices separately, and many devices need a way to recognize each other. Therefore, the router generates private IP addresses that are unique identifiers for each device that differentiate them on the network.
- **Public IP Address:** A public IP address, or external-facing IP address, is the primary address associated with a home or office network that connects business or home intranet network to their internet service provider (ISP). In most cases, this will be a router, and all devices that connect to the router communicates with other IP addresses using the router's IP address. Typically, ISPs have a large pool of IP addresses that they distribute to their customers. Public IP addresses are classified into two forms – dynamic and static.
  - **Dynamic IP Addresses:** A dynamic IP address is automatically assigned by ISPs to their customers from a large pool of IP addresses in their possession. Periodically, the ISPs re-assign IP addresses and put the older ones back into the pool to be used for other customers. Automating the regular movement of IP addresses means they don't have to carry out specific actions to re-establish a customer's IP address if they change location, for example. So, dynamic IP addresses change automatically and regularly.
  - **Static IP Addresses:** In contrast to dynamic IP addresses, static addresses remain consistent – they never change and they serve as a permanent internet address. Once the network assigns an IP address, it remains the same. Most individuals and businesses do not need a static IP address, but for businesses that plan to host their own server, it is crucial to have one. This is because a static IP address ensures that websites and email addresses tied to it will have a consistent IP address. Static IP Address provides information such as device location on the continent, which country, which city, and which Internet Service Provider provides internet connection to that particular device. Once, the ISP is known, the location of the device connected to the internet can traced.

3.1.3 Time Stamp

Time Stamp is a cryptographic digital attestation that a document or data was created or existed at a particular date and time, and has not been altered since a particular point in time, and serves as a trusted third party witnessing the existence and particulars of electronic data [21]. Timestamps are commonly used in computer systems to record and track when a piece of data or document was created, modified or accessed. By applying appropriate functions, developers can convert timestamps to formats like "YYYY-MM-DD HH:MM:SS" or other user-friendly representations.

3.1.4 QR Code

A quick response, better known as QR Code is a two-dimensional version of the barcode that stores information as a series of pixels in a square-shaped grid which can be read easily by a digital device. Unlike barcode's linear arrangement, QR codes can store much more data, because they're written both vertically and horizontally. They are frequently used to track information about products in a supply chain and are able to store up to 7089 digits or 4296 characters, including punctuation marks and special characters – the code can equally encode words and phrases such as website URLs, phone numbers etc [22]. QR codes are everywhere nowadays – from menus and boarding passes to payment links and product pages, QR codes help us access a range of services and information in the blink of an eye. It is possible to create QR codes in many different shapes and styles, but they all do the same job – they just look slightly different.

QR code Aztec code Maxi code PDF417 Semacode

3.1.5 Cryptographic Hash Function

A cryptographic hash function is a mathematical function that maps a bit string of arbitrary length into another compressed fixed-length bit string. The purpose of hash function is to produce a fingerprint of a file, message or other block of data and the values returned by a hash function are called message digest or simply hash values [23]. These functions are not just mere tools in cryptography, but are the backbone of digital integrity. A hash function in cryptography is fundamental in various applications, from securing sensitive data to ensuring the authenticity of digital communications. So, to be useful for message authentication, a hash function, h must have as a minimum, the following two properties [24]:

- h maps an input x of arbitrary finite bit-length, to an output $h(x)$ of fixed bit-length through the process of compression
- Given h and x, $h(x)$ is easy to compute.

In addition, hash functions that are relevant for cryptographic applications may fulfil one or several of the following requirements:

a) A hash function is pre-image resistant (or one-way) if for all pre-specified outputs, it is computationally difficult to find an input which hashes to that output, i.e., to find a preimage x' such that $h(x') = y$ when given any y for which a corresponding input is not known

b) A hash function is second-preimage resistant (or weak collision resistant) if it is computationally difficult to find any second input which has the same output as any specified input, i.e., given x, to find a second preimage $x' \neq x$ such that $h(x) = h(x')$.

c) A hash function is collision resistant (or strong collision resistant) if it is computationally difficult to find any distinct inputs x, x' that has the same output, i.e., $h(x) = h(x')$.

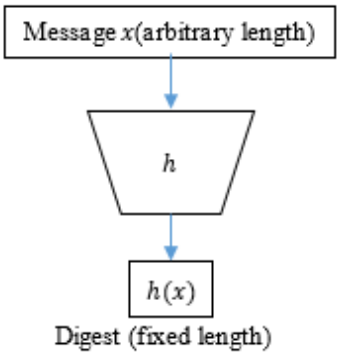

Fig 1. Cryptographic has function

Examples of cryptographic hash functions include MD2, MD4, MD5, SHA-1, RIPEMD or RIPEMD-160.

3.1.6 Public-Key Cryptography (Public and Private Key)

Public-key algorithms and structure are based on mathematical functions rather than on simple operations on bit patterns, such as are used in symmetric encryption algorithms [23]. In addition, public-key cryptography is asymmetric, involving the use of two separate keys in contrast to the symmetric conventional encryption, which uses only one secret key. Invariably, the private key is kept secret, but it is referred to as a private key rather than a secret key to avoid confusion with conventional symmetric encryption. The use of two keys has profound consequences in the areas of confidentiality, key distribution and authentication. As the name implies, the public key of the pair is made public for the general public to use, while the private key is known only to its owner. A general-purpose public-key cryptographic algorithm relies on one key for encryption and a different but related key for decryption. With this approach, all participants have access to the public keys, whereas the private keys are generated locally by each participant and therefore does not require to be distributed. So, as long as a user protects his or her private key, incoming communication is therefore secure. However, a user could change the private key at any time and publish the companion public key to replace the old public key.

A public-key scheme has the following properties:

- **Plaintext:** This is the readable message or data that is fed into the algorithm as input.
- **Encryption algorithm:** The encryption algorithm performs various transformation on the plaintext.
- **Public and private key:** This is a pair of keys that have been selected so that if one is used for encryption, the other is used for decryption. The exact transformations performed by the encryption algorithm depend on the public or private key that is provided as input.
- **Ciphertext:** This is the scrambled message produced as output. It depends on the plaintext and the key. For a given message, two different keys will produce two different ciphertexts.
- **Decryption algorithm:** The algorithm accepts the cipphertext and the matching key and produces the original plaintext.

II. DESIGN, METHODOLOGY AND DESCRIPTION OF THE NEW SCHEME

4.1 Design

a) **Source**

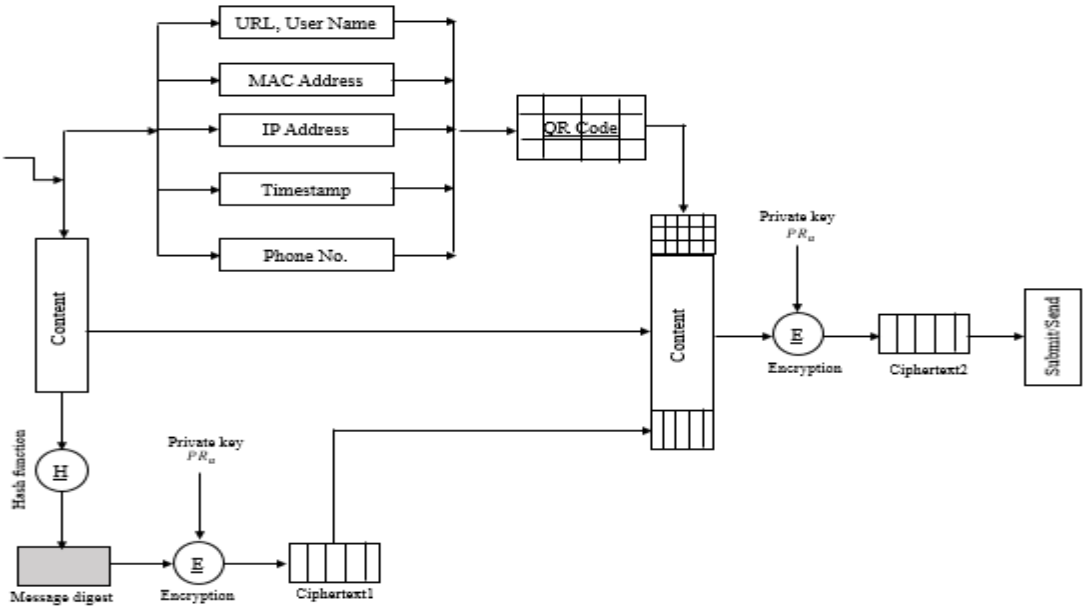

Fig 2. Scheme flow diagram (Source)

4.2 Methodology and Description

Figure 3.2 above, describes the source authentication mechanism as follows:

a) Source Design

Step 1: Here, the author prepares or upload the content (text, image, audio or video).

Step 2: In order to submit/send content, the scheme activates the author's authentication process which consists of two parts – content integrity assurance and author's identification. The purpose of the integrity assurance, is to ensure that the content or information is not tampered with as the information is being shared on the social media network. At this stage, the content or information is hashed using a cryptographic hash function such as SHA-1 to produce the message digest, followed by encryption of the digest using a private key (PR_a) to produce the ciphertext1. In addition, the source identification includes the acquisition of the author's identification credentials – the URL of the social media platform together with the user name, MAC address of the device, IP address, timestamp and phone number – this information is used to generate the QR code. The purpose of the QR code is first to ensure that the author's identification details are not displayed in plaintext, and secondly, to track automatically using the IP address.

Step 3: The trio of QR code, message/content and the ciphertext1 of the hashed message are further encrypted using the private-key to obtain the ciphertext2 of the trio and is appended to the content and submitted/sent accordingly.

b) Destination Design

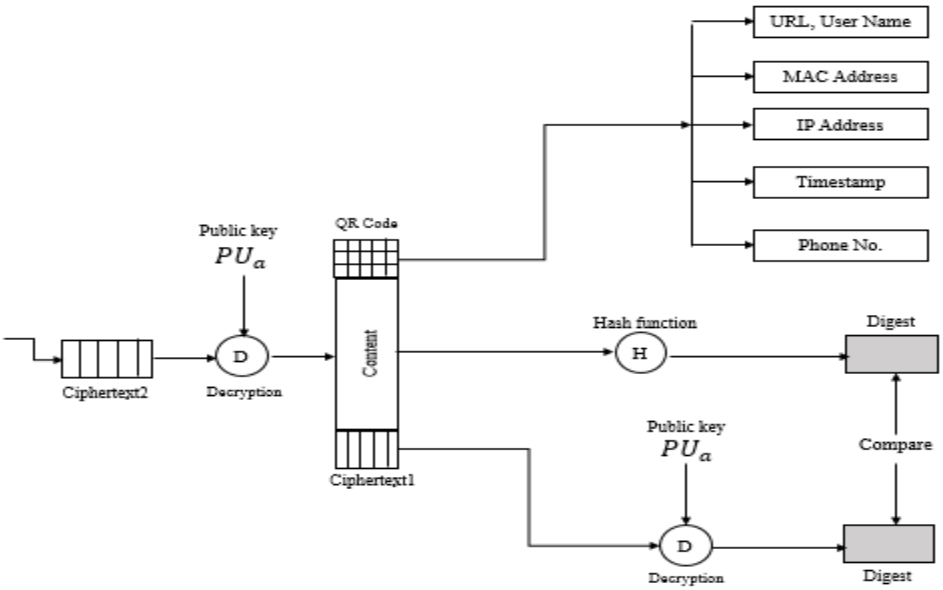

Fig 3. Content integrity test and reading of QR code

In Figure 3.3, the ciphertext2 containing the trio of QR code, message/content and ciphertext1 of the hashed message/content is decrypted using the public-key (PU_a) to unveil the trio. Next, the ciphertext1 containing the message/content is decrypted using the public key (PU_a) to obtain the digest while the message or content digest is obtained by hashing to produce its own digest which is then compared with the digest from the public-key decryption – if both digest remains the same, then the content has not been tampered with. Finally, the QR code used to store and conceal the identification details is then scanned to unveil the author's details and the author can be tracked and verified accordingly.

4.3 Online Privacy

I am sure that critics may raise the issue of privacy concern since one the identification data to be acquired from the author of content is phone number, but the phone number is not appended in plaintext, and so, it is when a message or content is harmful that the need to unmask the purveyor will become necessary. Just like Freedom of Information Act (FOIA) in Nigeria, if any person or group is desirous of accessing sensitive information, he/she is required to make a written application to the Legal and Regulatory Services (LRS) Department of the Commission responsible for responding to FOIA request. In likewise, any person or group interested in the identification details of the creator of an online content can request for the public-

key from a body that may be appointed as the key administrator. In this way, only authorised persons, groups or institution are allowed to have access to the author's details of the online content. This is to ensure the privacy of online users. In addition, this scheme is primarily designed to discourage or prevent/control the commission of the act and not necessarily to prosecute offenders, but in whichever case, it can as well be used to track and identify the actual source of the fake news whether it has been shared more than a million times or whether the post was created long ago.

III. Summary and Conclusion

Social media platforms offer a wealth of information, communication possibilities, and entertainment, however, inaccurate and misleading content is a persistent concern on online networks. The posting and sharing of misleading information or fake news, is done on the fly, but difficult to reverse or control, and worse still, the purveyor of such information might be difficult to identify or track; unlike traditional print or electronic media whose authors are well known and can be held responsible for any harm that the information may cause. With fake news spreading at the click of a button and at breakneck speed, tech companies are struggling to regulate it on social media, while grappling with public responsibility, definitions of free speech or freedom of expression, and identification of such content and its source. This research piece has provided another prism by which the excesses of the purveyors of disinformation, misinformation and mal-information, otherwise known as fake news, propaganda or misleading information can be curtailed. In the traditional electronic or print media, publishing or spreading fake news is a criminal offence in Nigeria and in most jurisdictions all over the world.

Chapter 33 of the Criminal Code in the Nigeria constitution provides for the criminalisation of various forms of defamation. For instance, Section 375 stipulates that "any person who publishes any defamatory matter is guilty of a misdemeanour and is liable to imprisonment for one year; and any person who publishes any defamatory matter knowing it to be false, is liable to imprisonment for two years." In the traditional media, it is easy to identify the purveyor of any news because the author's identification is well known to the general public but unlike in the social media platform, the authors of content is elusive and may not be known because the content or material does not contain any identification that could help to track the perpetuator; and even if eventually he/she is finally tracked, probably after the content may have been shared (say) a million times, the harm may have been done. So, the goal of this scheme is to discourage and prevent the creation and posting of harmful content on social media since such content will carry the identification of the author which is encapsulated in the QR code and secondly, the integrity of the content cannot be compromised as the cryptographic hash function will make any alteration an exercise in futility.

REFERENCES

[1] UNDP, "Guidance for the Implementation of UNDP iVerify," Tech for Democray and UNDP, 21 March 2023. [Online].

[2] National Academies, "Nobel Prize Summit Fuels Initiatives to Combat Misinformation and Disinformation and Build Trust in Science," National Academies, 22 June 2023. [Online].

[3] C. Ireton and J. Posetti, ***Journalism, 'Fake News' & Disinformation, Paris:*** UNESCO, 2018.

[4] UNPD - Europe and Central Asia, "Mapping and Analysis of Efforts to Counter Information Pollution in Europe and Central Asia Region," 22 December 2022. [Online].

[5] P. Salako, "Nigeria election triggers deluge of 'fake news' on social media," Al Jazeera, 15 February 2023. [Online]. Available: https://www.aljazeera.com/features/2023/2/15/nigeria-election-triggers-deluge-of-fake-news-on-social-media. [Accessed 12 February 2024].

[6] T. Hsu, S. A. Thompson and S. L. Mye, "Elections and Disinformation Are Colliding Like Never Before in 2024," The New York Times, 9 January 2024.

[7] E. Ajao, "AI, the 2024 U.S. election and the spread of disinformation," Techtarget, Enterprise AI, 2 February 2024. [Online]. Available: https://www.techtarget.com/searchenterpriseai/feature/AI-the-2024-US-election-and-the-spread-of-disinformation. [Accessed 12 February 2024].

[8] United Nations, "Countering Disinformation," United Nations, 12 August 2022. [Online]. Available: https://www.un.org/en/countering-disinformation. [Accessed 12 February 2024].

[9] Dictionary.com, "Information," Dictionary.com, [Online]. Available: https://www.dictionary.com/browse/information. [Accessed 12 February 2024].

[10] M. Ashikuzzaman, "Information: Types of information," Library & Information Science Community, 6 November 2023. [Online]. Available: https://www.lisedunetwork.com/definition-and-types-of-information/. [Accessed 13 February 2024].

[11] UNDP - Europe and Central Asia, "Rise Above: Countering Misinformation and Disinformation in Crisis Setting," UNDP, 22 December 2022. [Online]. Available: https://www.undp.org/eurasia/dis/misinformation. [Accessed 13 February 2024].

[12] American Psychological Association, "Using psychological science to fight misinformation: A guide for journalists," American Psychological Association, 29 November 2023. [Online]. Available: https://www.apa.org/topics/journalism-facts/psychology-misinformation-guide-journalists. [Accessed 15 February 2024].

[13] American Psychological Association, "Using psychology to understand and fight health misinformation," American Psychological Association, November 2023. [Online]. Available: https://www.apa.org/pubs/reports/health-misinformation. [Accessed 15 February 2024].

[14] P. Carpenter, "Get The 411 On Misinformation, Disinformation And Malinformation," Forbes, 13 January 2023. [Online]. Available: https://www.forbes.com/sites/forbesbusinesscouncil/2023/01/13/get-the-411-on-misinformation-disinformation-and-malinformation/?sh=3bf5cffe256a. [Accessed 16 February 2024].

[15] United Nations, "Understanding Hate Speech," United Nations, 2023. [Online]. Available: https://www.un.org/en/hate-speech/understanding-hate-speech/what-is-hate-speech. [Accessed 15 February 2024].

[16] W. M. Curtis, "Hate Speech," Encyclopedia Britannica, 18 January 2024. [Online]. Available: https://www.britannica.com/topic/hate-speech. [Accessed 15 February 2024].

[17] Council of Europe, "Dealing with propaganda, misinformation and fake news," Council of Europe, [Online]. Available: https://www.coe.int/en/web/campaign-free-to-speak-safe-to-learn/dealing-with-propaganda-misinformation-and-fake-news. [Accessed 17 February 2024].

[18] S. J. Dixon, "Misinformation on social media - Statistics & Facts," Statista, 10 January 2024. [Online]. Available: https://www.statista.com/topics/9713/misinformation-on-social-media/#topicOverview. [Accessed 17 February 2024].

[19] Meta, "Combating Misinformation," Meta, 2024. [Online]. Available: https://about.fb.com/news/tag/misinformation/. [Accessed 18 February 2024].

[20] USA Cyberspace Solarium Commission, "Countering Disinformation in the United States - CSC White Paper #6," USA Cyberspace Solarium Commission, 2021.

[21] Law Insider, "Time stamp definition," Law Insider, [Online]. Available: https://www.lawinsider.com/dictionary/time-stamp. [Accessed 22 February 2024].

[22] QR Code Generator, "QR Codes 101: A Beginner's Guide," QR Code Generator, [Online]. Available: https://www.qr-code-generator.com/qr-code-marketing/qr-codes-basics/. [Accessed 22 February 2024].

[23] W. Stallings, Network Security Essentials: Applications and Standards, Upper Saddle River: Pearson Education, Inc., 2007.

[24] R. Oppliger, Internet and Intranet Security, Norwood: Artech House, Inc., 2002.