United States Patent Application Publication	20250260720
Kind Code	A1
Publication Date	August 14, 2025
Inventor(s)	SHINDE; Sujit Raghunath et al.

METHOD AND SYSTEM FOR USABLE PHISHING PREVENTIVE INFORMATION SYSTEMS

Abstract

Method and system disclosed uses LLM with multi lingual support to helps identify phishing emails. The method creates sufficient visual indication for the user to notice it as an unusual email message. The method illustrates the email using relevant illustrations and/or emoticons in the notification bar of the phone or the desktop or web email client. Changes the font type face with kerning models as per the device form factor. By doing this disguised domain names or misspelled words can easily be noticed by the user. A sentiment analysis is utilized to identify if the email message is creating a sense of urgency for purported negative or positive event for the user. The result of sentiment analysis is then used to also recommend 2 factor authentication. Thus allow user to establish credibility of the email. accordingly the email header and contents are updated appropriately.

Inventors: **SHINDE; Sujit Raghunath (Thane, IN), DOKE; Pankaj Harish (Mumbai, IN), BHAVSAR; Karan Rajesh (Thane, IN), KIMBAHUNE; Sanjay Madhukar (Mumbai, IN)**

Applicant: **Tata Consultancy Services Limited** (Mumbai, IN)

Family ID: **94480954**

Assignee: **Tata Consultancy Services Limited (Mumbai, IN)**

Appl. No.: **19/044967**

Filed: **February 04, 2025**

Foreign Application Priority Data

IN	202421008938	Feb. 09, 2024

Publication Classification

Int. Cl.: **H04L9/40** (20220101); **H04L41/16** (20220101)

U.S. Cl.:

CPC **H04L63/1483** (20130101); **H04L41/16** (20130101); **H04L63/0853** (20130101);

Background/Summary

PRIORITY CLAIM
[0001] This U.S. patent application claims priority under 35 U.S.C. § 119 to: Indian Patent Application number 202421008938, filed on Feb. 9, 2024. The entire contents of the aforementioned application are incorporated herein by reference.
TECHNICAL FIELD
[0002] The embodiments herein generally relate to the field of phishing prevention in information systems and, more particularly, to a method and system for providing usable phishing preventive information systems.
BACKGROUND
[0003] Ease of internet access, smart phone availability has brought in exponential growth in communication over Internet Protocol (IP) based services. With large volumes of message exchanges via Internet Protocol (IP) based messaging, phishing attacks have penetrated the digital communication, taking advantage of vulnerabilities that the information technology system poses to users. Even though security techniques against phishing attacks have been developed and explored, users still fall prey to phishing attacks, specifically targeted through messages. This is mainly because end users are tricked to click bad links, or open infected attachments. “Time pressure”, “work delivery pressure” and so on tend to put the user in speed reading mode. The intruders grab this opportunity where the user easily can misjudge the received message as an authentic message due to intentionally created name similarity between known and unknown sender friendly domain name. These, so called similar senders may trick around spelling variations, or character shapes and the like, which a user may fail to notice. Thus, to address the above concerns, one important aspect is to understand on human machine interaction challenges to provide effective solutions.
[0004] Various approaches are explored to identify such malicious senders sending phishing messages and warn or alert users (message receivers or viewers) against such messages. Senders Identity (ID) validations via Sender Profile Framework (SPF) Check, cross verification using blacklisted domain names (static list) are commonly used approaches to identify malicious sender. However, the attackers are ahead to find loopholes to get through these standard check mechanisms. Further, there are few works proposing various possible approaches that help to better identify phishing attack and indicate via audio visual phishing warning to user. However the technical solutions to rightly achieve them are not explored. Further, phishing preventive solutions generally are developed with the assumption that the user is technical literate. However, large population of developing countries that uses smart phones for communication is less literate, also referred to as Basic Emergent Users (BEUs). These BEUs have only a basic knowledge of accessing smart phones and applications on them. Limited knowledge of commonly used languages like English can be a potential loophole that provides scope to the attackers to easily carry out the phishing activities.
[0005] Exploring robust techniques for phishing attack detection with easily noticeable, interpretable, effective alert indication to the end user (receiver) is an open area of research. Furthermore, personalization in alert indication for effectively communicating an alert is critical

considering variation is technological literacy levels of end users.
[0006] Usable security is a branch of security that focuses on preventing threats to user security and privacy that arise from the interaction of humans (users) with computer systems. Unlike traditional system and network security, it focuses on users, analyzing their behavior, mental models and decision-making processes and incorporate it in during security design. Thus, needed are robust usable security designs for information systems that collect, store, process and transmit data and digital information.

SUMMARY

[0007] Embodiments of the present disclosure present technological improvements as solutions to one or more of the above-mentioned technical problems recognized by the inventors in conventional systems.
[0008] For example, in one embodiment, a method for usable phishing prevention is provided. The method includes initiating, for each message among a plurality of messages received from one or more senders associated with sender ID and geo-sensor equipped sender devices, a Sender Profile Framework (SPF) check by checking records storing authorized domain names, via a plugin usable security module executed by a message exchange and communication server, wherein the sender ID tagged as SPF-fail indicates that a domain name associated with the sender ID belongs to an unauthorized domain IP address, wherein a SPF stores legal record of a plurality of domains and associated domain owner and organization details.
[0009] Further, the method includes modifying the records with a set of custom headers comprising a plurality of parameters associated with a sender device profile, wherein the plurality of parameters comprising geocoding, device type, device type profile, sensor type, sensor type profile and accurate real-time positioning and timing services received from satellite agencies.
[0010] Further, the method includes analyzing, the one or more sender IDs tagged with SPF-fail in conjunction with one or more of the set of custom headers by performing one of: [0011] (i) initiating a two factor authentication (2FA), if the message is associated with at least one of a mass mailer communication and an unauthorized domain IP address, the 2FA comprises: communicating with a domain owner, via a voice bot executing a Voice User Interface (VUI), based on information extracted from the contact details to confirm authorization of the unauthorized domain IP address, wherein the domain is classified as rogue if authorization is not confirmed; identifying the sender associated with rogue domain as malicious and quarantining the message; and communicating a legal notice to the domain owner via Voice user interfaces (VUIs) and over an email, wherein the legal notice is obtained from a legal expert using voice to text technique; [0012] (ii) stripping body of the message, if the message does not belong to the mass mailer communication and does not have suspected domain name, and generating a checksum to derive presence of the mass mailer communication using checksum hit technique, wherein if the hits are beyond a threshold count obtain the legal notice using text to voice technique and communicate to the the sender is identified as potentially malicious; and [0013] (iii) if the domain IP address is detected to be mobile based on the geo code associated with the sender device present in the set of custom headers and captured in message header, then identify the sender as potentially malicious. Further, the method includes processing the message using
[0014] prompt engineering, by a first LLM in conjunction with a phishy-domain-permutations database, if the sender of the message is identified as potentially malicious, wherein the processing comprises: [0015] (i) analyzing the message body to provide reasons for sender indicated as malicious, wherein the reasons are in natural language and machine interpretable language; [0016] (ii) iteratively processing the reasons to generate short reasons excluding technical jargons to be embedded in the body of the message; [0017] (iii) generating an image using text-to-Image models, wherein the image is indicative of potentially malicious mail and sentiment in the mail; and [0018] (iv) including a warning indicating IP address is mobile and not fixed.
[0019] Furthermore, the method includes modifying the Document Object Module (DOM) instance

for the message to incorporate the processed results of the LLM, wherein the message is displayed to a receiver on a receiver device with one or more phishing warnings, wherein the phishing warnings are displayed at the DOM nodes based on positions indicated by the LLM, and wherein phishing warnings in form of text message are displayed with change the font type face with kerning models as per a device form factor of a receiver device.

[0020] In another aspect, a system for usable phishing prevention is provided. The system comprises a memory storing instructions; one or more Input/Output (I/O) interfaces; and one or more hardware processors coupled to the memory via the one or more I/O interfaces, wherein the one or more hardware processors are configured by the instructions to initiate, for each message among a plurality of messages received from one or more senders associated with sender ID and geo-sensor equipped sender devices, a Sender Profile Framework (SPF) check by checking records storing authorized domain names, via a plugin usable security module executed by a message exchange and communication server, wherein the sender ID tagged as SPF-fail indicates that a domain name associated with the sender ID belongs to an unauthorized domain IP address, wherein a SPF stores legal record of a plurality of domains and associated domain owner and organization details.

[0021] Further, the one or more hardware processors are configured to modify the records with a set of custom headers comprising a plurality of parameters associated with a sender device profile, wherein the plurality of parameters comprising geocoding, device type, device type profile, sensor type, sensor type profile and accurate real-time positioning and timing services received from satellite agencies.

[0022] Further, the one or more hardware processors are configured to analyze, the one or more sender IDs tagged with SPF-fail in conjunction with one or more of the set of custom headers by performing one of: [0023] (i) initiating a two factor authentication (2FA), if the message is associated with at least one of a mass mailer communication and an unauthorized domain IP address, the 2FA comprises: communicating with a domain owner, via a voice bot executing a Voice User Interface (VUI), based on information extracted from the contact details to confirm authorization of the unauthorized domain IP address, wherein the domain is classified as rogue if authorization is not confirmed; identifying the sender associated with rogue domain as malicious and quarantining the message; and communicating a legal notice to the domain owner via Voice user interfaces (VUIs) and over an email, wherein the legal notice is obtained from a legal expert using voice to text technique; [0024] (ii) stripping body of the message, if the message does not belong to the mass mailer communication and does not have suspected domain name, and generating a checksum to derive presence of the mass mailer communication using checksum hit technique, wherein if the hits are beyond a threshold count obtain the legal notice using text to voice technique and communicate to the the sender is identified as potentially malicious; and [0025] (iii) if the domain IP address is detected to be mobile based on the geo code associated with the sender device present in the set of custom headers and captured in message header, then identify the sender as potentially malicious.

[0026] Further, the one or more hardware processors are configured to process the message using prompt engineering, by a first LLM in conjunction with a phishy-domain-permutations database, if the sender of the message is identified as potentially malicious, wherein the processing comprises: [0027] (i) analyzing the message body to provide reasons for sender indicated as malicious, wherein the reasons are in natural language and machine interpretable language; [0028] (ii) iteratively processing the reasons to generate short reasons excluding technical jargons to be embedded in the body of the message; [0029] (iii) generating an image using text-to-Image models, wherein the image is indicative of potentially malicious mail and sentiment in the mail; and [0030] (iv) including a warning indicating IP address is mobile and not fixed.

[0031] Furthermore, the one or more hardware processors are configured to modify the Document Object Module (DOM) instance for the message to incorporate the processed results of the LLM,

wherein the message is displayed to a receiver on a receiver device with one or more phishing warnings, wherein the phishing warnings are displayed at the DOM nodes based on positions indicated by the LLM, and wherein phishing warnings in form of text message are displayed with change the font type face with kerning models as per a device form factor of a receiver device.

[0032] In yet another aspect, there are provided one or more non-transitory machine-readable information storage mediums comprising one or more instructions, which when executed by one or more hardware processors causes a method for usable phishing prevention. The method includes initiating, for each message among a plurality of messages received from one or more senders associated with sender ID and geo-sensor equipped sender devices, a Sender Profile Framework (SPF) check by checking records storing authorized domain names, via a plugin usable security module executed by a message exchange and communication server, wherein the sender ID tagged as SPF-fail indicates that a domain name associated with the sender ID belongs to an unauthorized domain IP address, wherein a SPF stores legal record of a plurality of domains and associated domain owner and organization details.

[0033] Further, the method includes modifying the records with a set of custom headers comprising a plurality of parameters associated with a sender device profile, wherein the plurality of parameters comprising geocoding, device type, device type profile, sensor type, sensor type profile and accurate real-time positioning and timing services received from satellite agencies.

[0034] Further, the method includes analyzing, the one or more sender IDs tagged with SPF-fail in conjunction with one or more of the set of custom headers by performing one of: [0035] (i) initiating a two factor authentication (2FA), if the message is associated with at least one of a mass mailer communication and an unauthorized domain IP address, the 2FA comprises: communicating with a domain owner, via a voice bot executing a Voice User Interface (VUI), based on information extracted from the contact details to confirm authorization of the unauthorized domain IP address, wherein the domain is classified as rogue if authorization is not confirmed; identifying the sender associated with rogue domain as malicious and quarantining the message; and communicating a legal notice to the domain owner via Voice user interfaces (VUIs) and over an email, wherein the legal notice is obtained from a legal expert using voice to text technique; [0036] (ii) stripping body of the message, if the message does not belong to the mass mailer communication and does not have suspected domain name, and generating a checksum to derive presence of the mass mailer communication using checksum hit technique, wherein if the hits are beyond a threshold count obtain the legal notice using text to voice technique and communicate to the the sender is identified as potentially malicious; and [0037] (iii) if the domain IP address is detected to be mobile based on the geo code associated with the sender device present in the set of custom headers and captured in message header, then identify the sender as potentially malicious.

[0038] Further, the method includes processing the message using prompt engineering, by a first LLM in conjunction with a phishy-domain-permutations database, if the sender of the message is identified as potentially malicious, wherein the processing comprises: [0039] (i) analyzing the message body to provide reasons for sender indicated as malicious, wherein the reasons are in natural language and machine interpretable language; [0040] (ii) iteratively processing the reasons to generate short reasons excluding technical jargons to be embedded in the body of the message; [0041] (iii) generating an image using text-to-Image models, wherein the image is indicative of potentially malicious mail and sentiment in the mail; and [0042] (iv) including a warning indicating IP address is mobile and not fixed.

[0043] Furthermore, the method includes modifying the Document Object Module (DOM) instance for the message to incorporate the processed results of the LLM, wherein the message is displayed to a receiver on a receiver device with one or more phishing warnings, wherein the phishing warnings are displayed at the DOM nodes based on positions indicated by the LLM, and wherein phishing warnings in form of text message are displayed with change the font type face with kerning models as per a device form factor of a receiver device.

[0044] It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the invention, as claimed.

Description

BRIEF DESCRIPTION OF THE DRAWINGS

[0045] The accompanying drawings, which are incorporated in and constitute a part of this disclosure, illustrate exemplary embodiments and, together with the description, serve to explain the disclosed principles:

[0046] FIGS. **1**A and **1**B illustrates a system for usable phishing prevention in an Internet Protocol (IP) messaging environment, in accordance with some embodiments of the present disclosure.

[0047] FIG. **1**C is an architectural overview of the system for usable phishing prevention, in accordance with some embodiments of the present disclosure.

[0048] FIG. **2** is a functional diagram of the system for usable phishing prevention, in accordance with some embodiments of the present disclosure.

[0049] FIGS. **3**A and **3**B (collectively referred as FIG. **3**) is a flow diagram illustrating a method for usable phishing prevention, using the system depicted in FIGS. **2**A and **2**B, in accordance with some embodiments of the present disclosure.

[0050] FIGS. **4**A through **7**F explain stage by stage processing flow by the system for a sample email (Simple Mail Transfer Protocol-SMTP) received from a sender, wherein the system modifies the sample message to indicate phishing warning before sending it to a receiver, in accordance with some embodiments of the present disclosure.

[0051] FIGS. **8**A and **8**B are example interactive User Interface (UIs) with phishing warnings displayed to a receiver on receiver's device by the system for different types of phishing attack via potentially malicious emails detected by the system, in accordance with some embodiments of the present disclosure.

[0052] FIGS. **9**A and **9**B are example interactive User Interface (UIs) with warnings displayed to a receiver on receiver's device by the system for different types of phishing attack via potentially malicious chat messages detected by the system, in accordance with some embodiments of the present disclosure.

[0053] FIGS. **10**A through **10**D explain stage by stage processing flow by the system for a sample message (instant chat based on Extensible Messaging Presence Protocol-XMPP) received from a sender, wherein the system modifies the sample message to indicate phishing warning before sending it to a receiver, in accordance with some embodiments of the present disclosure.

[0054] It should be appreciated by those skilled in the art that any block diagrams herein represent conceptual views of illustrative systems and devices embodying the principles of the present subject matter. Similarly, it will be appreciated that any flow charts, flow diagrams, and the like represent various processes which may be substantially represented in computer readable medium and so executed by a computer or processor, whether or not such computer or processor is explicitly shown.

DETAILED DESCRIPTION

[0055] Exemplary embodiments are described with reference to the accompanying drawings. In the figures, the left-most digit(s) of a reference number identifies the figure in which the reference number first appears. Wherever convenient, the same reference numbers are used throughout the drawings to refer to the same or like parts. While examples and features of disclosed principles are described herein, modifications, adaptations, and other implementations are possible without departing from the scope of the disclosed embodiments.

[0056] Embodiments of the present disclosure provide a method and system for usable phishing preventive information systems. Information system is an integrated set of components for

collecting, storing, and processing data and for providing information, knowledge, and digital products. Mobile phones (smart phones) carried by end users, the servers and the like are components of information systems involved in message communication. Message communication in information system is mainly based on the Internet Protocol (IP), for example a Simple Mail Transfer Protocol (SMTP) for email communication, Extensible Messaging and a Presence Protocol (XMPP) for instant messaging (IM) and the like used by chat applications.

[0057] While reading a message, for example the email, the end user (receiver) if is focused on reading the message, he or she may read the email thoroughly well to easily spot a maligned email. However, most of the time in a hurry or when distracted he may be in speed reading mode and completely overlooks minute signs that could help suspect the email. While in speed reading mode, the brain uses it is subconscious visual memory to recognize words by their shape, size and structure and ignore minute spelling errors. An intruder, also referred to as sender, exploits this human attribute to replace visually similar characters with bait domains and email-ids (sender identities (IDs)). The method and system disclosed herein analyzes and highlights suspicious indicators to the user to not visually overlook them while speed reading the emails. The highlights refer to phishing warnings are displayed based on suggestion by Large Language Model (LLM) so as to be in accordance with Usability Principles of Communications that require proximity of warning based on Gestalt principles. the ability of intent extraction using LLM is a recent development. People skilled in this aspect work in the domain of NLP and AI ML and Interaction Design.

[0058] Messaging based communication is a infrastructure services skillset. Works in this domain are focused on issues pertaining to communications and networks. Similarly technical focus in information security are primarily related to security aspects related to computing elements. Whereas works in the design domain focus only on visual aesthetics and visual design. Thus, a person with security background is unlikely to have design or LLM awareness to a larger extent, thus the method combines the LLM capabilities to security domain.

[0059] In some scenarios the suspicious emails may be legitimate emails and not intended to be blocked or quarantined, for example with non-deliberate spelling errors, the method highlights such indicators to the user and provide a medium to allow user to ascertain email's authenticity, by authenticating either sender's domain or embedded Uniform Resource Locator (URL) domains. Often such emails are sent in bulk and there could be several recipients receiving similar email with same content with changed salutation to the respective recipient. Hence, the method optimizes the two factor authentication (2FA) by caching (recording) the email content HASH and annotating it with the outcome of 2FA authentication. This approach enables to apply 2FA only once among all the recipients, reducing the processing overhead.

[0060] Referring now to the drawings, and more particularly to FIGS. **1**A through **10**D, where similar reference characters denote corresponding features consistently throughout the figures, there are shown preferred embodiments, and these embodiments are described in the context of the following exemplary system and/or method.

[0061] FIGS. **1**A and **1**B illustrates a system **102** for usable phishing prevention in an Internet Protocol (IP) messaging environment **100**, in accordance with some embodiments of the present disclosure. FIG. **1**C is an architectural overview of the system **102** for usable phishing prevention, in accordance with some embodiments of the present disclosure.

[0062] The system **102** herein includes a message exchange server executing a plugin usable security module while transferring messages between a sender and receiver enabled by communication channel set up between the device **104** (sender end device) to a device **106** (receiver end device). The message exchange server can for example be a SMTP server, an IM server, and the like. The system and the method flow is explained herein with email as an example of IP based messaging and does not indicate a limitation of the IP based messaging types of the system **102** can handle. FIGS. **8**A and **8**B are example interactive User Interface (UIs) with

phishing warnings displayed to a receiver on receiver's device by the system for different types of phishing attack via potentially malicious emails detected by the system, in accordance with some embodiments of the present disclosure. FIGS. **9**A and **9**B are example interactive User Interface (UIs) with warnings displayed to a receiver on receiver's device by the system for different types of phishing attack via potentially malicious chat messages detected by the system, in accordance with some embodiments of the present disclosure.

[0063] Phishing emails usually have spelling/grammatical errors, and or use incorrect or disguised domain names. The system **102** herein utilizes LLM with multilingual support to helps identify such phishing emails. For any such emails, the system **102** creates sufficient visual indication as seen in FIGS. **8**A and **8**B when the email (message) is read by a receiver (user) on his/her device **106**. The system **102**. [0064] 1. illustrates the email using relevant illustrations and/or emoticons in the notification bar of the phone or the desktop or web email client. [0065] 2. Change the font type face with kerning models as per the device form factor. By doing this disguised domain names or misspelled words can easily be noticed by the user. The technique to change font type face is as explained in applicant's granted patent titled “Method and device for dynamic viewport generation to enhance viewport usability” with U.S. Pat. No. 11,132,495B1 patent number at USPTO and not explained herein for brevity. [0066] 3. Uses sentiment analysis to identify if the email message is creating a sense of urgency for purported negative or positive event for the user. [0067] 4. Use sentiment analysis to also recommend 2FA based authentication. Thus allow user to establish credibility of the email.

[0068] In accordance with the above operations the system **102** updates email header and contents, to be displayed to user on device **106** so as to be distinguishably noticed by a reader (receiver).

[0069] The existing messaging systems do not modify the email (message) content, with visual illustration or with local language warnings highlighting false domains of phishing traps for the less literate users. The existing systems do not use generative Artificial Intelligence (such as the LLM used herein) for generating variations on authentic domains. Also the existing systems do not provide any 2FA (explained in detail later) for validating the credibility of the email sender for Business to Business (B2B) as well as Business to Consumer (B2C).

[0070] For existing systems, the technical challenge in inserting warnings lies in finding the appropriate Document Object Module (DOM) node for insertion of “Warning” nodes in a pre-formatted phishing email for maximum effectiveness from user behavior point of view considering in accordance with Usability Principles of Communications that require proximity of warning based on Gestalt principles, which can be understood from information website https://digital.gov/communities/user-experience/#related-resources-2 Also, definitions for usability can be further understood in accordance with information provided at https://www.iso.org/obp/ui/#iso: std: iso: 9241:-11: ed-2: v1: en.

[0071] It is critical for business to understand given a DOM what content in a phishing email should be added to effect behavior change in recipient of the email, and during technical implementation determine minimal number of nodes of the DOM for intended updating of the message.

[0072] Geo tracing of servers (sender device) for multi-factor authentication is used by the system. Currently, the scope of authentication is limited to cryptography based approaches such as certificates and CA. However, geotagging of servers as a multi-factor for authentication for emails is not currently in the State-of-the-Art (SOTA) approaches. Servers can be made mobile and mobile devices can be made servers. Multi-factor authentication of such scenarios is not present in the current SOTA. Servers can either be in Datacenter's or on Moving Vehicle.

[0073] As can be seen in FIGS. **8**A through **9**B example UI illustrations as displayed by system on user end devices (device **106**) when user (receiver) reads the messages (email or chat messages form IM applications): The system provides: [0074] 1. Embedding to Illustration (Embedding2Illustration): Illustrates the email using relevant illustrations and/or emoticons in the

notification bar of the phone or the desktop or web email client. Alternate Text and or audio is generated for accessibility towards Visual Impaired Users. Adaptable as it can be photographs instead of icons) [0075] 2. Embedding to Web Domain (Embedding2Webdomain): Phishing emails usually have spelling/grammatical errors, and or use incorrect or disguised domain names. Embodiments herein use LLM to help identify such phishing emails. [0076] 3. Embedding to Typeface (Embedding2TypeFace) with Kerning Models per Form Factor: Change the font type face with kerning models as per the device form factor. By doing this disguised domain names or misspelled words can easily be noticed by the user. [0077] 4. Embedding to Sentiment Analysis (Embedding2SentimentAnalysis): Uses sentiment analysis to identify if the email message is creating a sense of urgency for purported negative or positive event for the user. [0078] 5. Automated2FA: Use sentiment analysis to also recommend 2 factor authentication. Thus, allow users to establish credibility of the email.

[0079] The registry has contact [email, phone] of authorized person. As can be seen in FIG. **8**A through **9**B, for 2FA, the email (chat message) is updated to include a “validate” tap/click for automated 2FA on tapping it, an auto callback is generated to the registered number—where in a special key say hash, pound is pressed—this validates the email as authentic on the authorized person (domain owner) side, an automated IVR is setup—which picks up a call and send a hash or pound key automatically—no human is required this email, if required, is auto-authenticated from the domain of the sender organization as is known, this system and number is not with intruder—hence 2FA will fail—so—the message will not be validated. This can be automated, or manual based on the budget of the receiver of the email who can accordingly initiate 2FA.

[0080] This architecture depicted in FIGS. **1**A and **1**B can be applied to master-slave systems and also to peer-to-peer systems. The machine/system in the block diagram is a role which can be performed by the same or different devices.

[0081] FIG. **2** is a functional diagram of the system **102** for usable phishing prevention, in accordance with some embodiments of the present disclosure. In an embodiment, the system **102** includes a processor(s) **204**, communication interface device(s), alternatively referred as input/output (I/O) interface(s) **206**, and one or more data storage devices or a memory **202** operatively coupled to the processor(s) **204**. The system **102** with one or more hardware processors is configured to execute functions of one or more functional blocks of the system **102**.

[0082] Referring to the components of system **102**, in an embodiment, the processor(s) **204**, can be one or more hardware processors **204**. In an embodiment, the one or more hardware processors **204** can be implemented as one or more microprocessors, microcomputers, microcontrollers, digital signal processors, central processing units, state machines, logic circuitries, and/or any devices that manipulate signals based on operational instructions. Among other capabilities, the one or more hardware processors **204** are configured to fetch and execute computer-readable instructions stored in the memory **202**. In an embodiment, the system **102** can be implemented in a variety of computing systems including laptop computers, notebooks, hand-held devices such as mobile phones, workstations, mainframe computers, servers, and the like.

[0083] The I/O interface(s) **206** can include a variety of software and hardware interfaces, for example, a web interface, a graphical user interface and the like and can facilitate multiple communications within a wide variety of networks N/W and protocol types, including wired networks, for example, LAN, cable, etc., and wireless networks, such as WLAN, cellular and the like. In an embodiment, the I/O interface(s) **206** can include one or more ports for connecting to a number of external devices or to another server or devices such as multiple sender devices (device **104**) and multiple receiver devices such as device (**106**). Further, a GPS sensor is attached the message exchange and communication server via the I/O port (**206**)

[0084] The memory **202** may include any computer-readable medium known in the art including, for example, volatile memory, such as static random access memory (SRAM) and dynamic random access memory (DRAM), and/or non-volatile memory, such as read only memory (ROM), erasable

programmable ROM, flash memories, hard disks, optical disks, and magnetic tapes.
[0085] In an embodiment, the memory **202** includes a plurality of modules **210** such as the usable security module (implemented as a plugin) that provides a 2FA module (not shown in FIG. **2**), a first LLM (not shown) for message body modification functionality, a second LLM not shown) for generating phishy-domain-permutations database.
[0086] The plurality of modules **210** include programs or coded instructions that supplement applications or functions performed by the system **102** for executing different steps involved in the process of usable phishing prevention in information system. The plurality of modules **210**, amongst other things, can include routines, programs, objects, components, and data structures, which performs particular tasks or implement particular abstract data types. The plurality of modules **210** may also be used as, signal processor(s), node machine(s), logic circuitries, and/or any other device or component that manipulates signals based on operational instructions. Further, the plurality of modules **210** can be used by hardware, by computer-readable instructions executed by the one or more hardware processors **204**, or by a combination thereof. The plurality of modules **210** can include various sub-modules (not shown).
[0087] Further, the memory **202** may comprise information pertaining to input(s)/output(s) of each step performed by the processor(s) **204** of the system **102** and methods of the present disclosure.
[0088] Further, the memory **202** includes a database **208**. The database (or repository) **208** may include a plurality of abstracted pieces of code for refinement and data that is processed, received, or generated as a result of the execution of the plurality of modules in the module(s) **210**. The database also includes the phishy-domain-permutations database, the personalized threat profile for each receiver.
[0089] Although the database **208** is shown internal to the system **102**, it will be noted that, in alternate embodiments, the database **208** can also be implemented external to the system **102**, and communicatively coupled to the system **102**. The data contained within such an external database may be periodically updated. For example, new data may be added into the database (not shown in FIG. **2**) and/or existing data may be modified and/or non-useful data may be deleted from the database. In one example, the data may be stored in an external system, such as a Lightweight Directory Access Protocol (LDAP) directory and a Relational Database Management System (RDBMS). Functions of the components of the system **102** are now explained with reference to FIG. **3** through FIG. **9**B.
[0090] FIGS. **3**A and **3**B (collectively referred as FIG. **3**) is a flow diagram illustrating a method **300** for usable phishing prevention, using the system depicted in FIGS. **2**A and **2**B, in accordance with some embodiments of the present disclosure.
[0091] In an embodiment, the system **102** comprises one or more data storage devices or the memory **202** operatively coupled to the processor(s) **204** and is configured to store instructions for execution of steps of the method **300** by the processor(s) or one or more hardware processors **104**. The steps of the method **300** of the present disclosure will now be explained with reference to the components or blocks of the system **102** as depicted in FIGS. **1**A and **1**B, the steps of flow diagram as depicted in FIG. **3** and example of email (message) processing as depicted in FIGS. **4**A through **7**F. Although process steps, method steps, techniques or the like may be described in a sequential order, such processes, methods, and techniques may be configured to work in alternate orders. In other words, any sequence or order of steps that may be described does not necessarily indicate a requirement that the steps be performed in that order. The steps of processes described herein may be performed in any order practical. Further, some steps may be performed simultaneously.
[0092] Referring to the steps of the method **300**, at step **302** of the method **300**, the one or more hardware processors **204** of the message exchange and communication server **102** (system **102**) by executing the plug-in usable security module are configured by the instructions to initiate, for each message among a plurality of messages received from one or more senders a Sender Profile Framework (SPF) check by checking records storing authorized domain names. The SPF check is a

standard process used in the SMTP. An SPF record identifies the mail servers and domains that are allowed to send email on behalf of your domain. Receiving servers check your SPF record to verify that incoming messages that appear to be from your organization are sent from servers allowed by you. Domains can have one SPF record.

[0093] The example message type (email) as received is depicted in FIG. **4**B (stage **0**) process step of the method **300**. As can be understood email is an example message type and the method **300** is equally applicable for IM type of messages using XMPP protocol for instant chat applications as depicted in FIGS. **9**A and **9**B.

[0094] Each sender is associated with sender ID and geo-sensor equipped sender devices. The SPF check is depicted in detail in FIG. **5**A. (Stage **1**-step **1**.**1**) wherein the sender ID tagged as SPF-fail indicates that a domain name associated with the sender ID belongs to an unauthorized domain IP address, wherein a SPF stores legal record of a plurality of domains and associated domain owner and organization details.

[0095] At step **304** of the method **300**, the one or more hardware processors **204** of the message exchange and communication server **102** (system **102**) by executing the plug-in usable security module are configured by the instructions to modify the records (for example Mail Exchange (MX) records if the messages are emails) with a set of custom headers comprising a plurality of parameters associated with a sender device profile. As depicted in MXRecordDTD.XML in FIG. **5**B (stage **1**-step **1**.**1**-**2**) and **5**C (stage **1**-step **1**.**1**-**3**) and flow diagram of FIG. **5**M, the set of custom headers (also referred to as custom headers or custom fields in message header) are shown in italics and corresponding to the plurality of parameters comprising geocoding, device type, device type profile, sensor type, sensor type profile and accurate real-time positioning and timing services received from satellite agencies.

[0096] Generally, all device manufacturers give a device profile (herein referred to as sender device profile). For example, in case of Android™ phones, a logon to the developer portal of Google™ enables access to all Android world over which device details are provided, as depicted in example of Table 1A and 1B.

TABLE-US-00001 TABLE 1 Model RAM Form System on Brand Device Manufacturer Name (TotalMem) Factor Chip samsung m31 Samsung Galaxy 5843MB Phone Samsung M31 Exynos 9611 samsung j6lte Samsung Galaxy 2948MB Phone Samsung J6 Exynos 7870 samsung j7elte Samsung Galaxy 1437MB Phone Samsung J7 Exynos 7580 samsung a02q Samsung Galaxy 3741MB Phone Qualcomm A02s SDM450 Android Open User- User- Screen Screen SDK GL ES Install perceived perceived GPU Sizes Densities ABIs Versions Versions base ANR rate crash rate ARM Mali 1080 × 420 arm64-v8a; 31 196610 2 0.00% 0.00% G72 2340 armeabi; (1050 MHz) armeabi-v7a ARM Mali 720 × 320 armeabi; 29 196610 2 0.00% 0.00% T830 1480 armeabi-v7a (700 MHz) ARM Mali 720 × 320 armeabi; 23 196609 1 0.00% 0.00% T720 1280 armeabi-v7a (668 MHz) Qualcom 720 × 280 armeabi; 31 196610 1 0.00% 0.00% m Adreno 1600 armeabi-v7a 506 (600 MHz) ARM Mali 1080 × 420 arm64-v8a; 29 196610 1 0.00% 0.00% G71 2220 armeabi; (1100 MHz) armeabi-v7a

[0097] In general devices use GPS for location tracking. However, the system **102** herein also provides use of NavIC™ system offered by satellite agency such as Indian Space Research Organization (ISRO). Thus, on-device sensor or GPS sensor attached as an external sensor to the device/machine/system **102** works with satellites constellations like GPS, NavIC™ to get device latitude and longitude (latOlong).

[0098] At step **306** of the method **300**, the one or more hardware processors **204** of the message exchange and communication server **102** (system **102**) by executing the plug-in usable security module are configured by the instructions to analyze, the one or more sender IDs tagged with SPF-fail in conjunction with one or more of the set of custom headers to check. The analysis is performed using at least one of the following steps as below (also depicted in flow diagram of FIG. **5**G-**1**): [0099] (i) Initiate a two factor authentication (2FA) as depicted in stage FIG. **5**D (stage**1**-

step **1**.**2**) if the message is associated with at least one of a mass mailer communication and an unauthorized domain IP address.
[0100] There are well known approaches used to detect a mass mailer such as detecting one of: [0101] One sender domain—1 IP and many receivers. [0102] One sender domain—N IP and many receivers. [0103] (Above characterized by timestamps been in vicinity or same)
[0104] The 2FA comprises (also depicted in flow diagram of FIG. **5**G): [0105] Communicating with a domain owner, via a voice bot executing a Voice User Interface (VUI), based on information extracted from the contact details to confirm authorization of the unauthorized domain IP address. For example for the email service this contact information is obtained from a Whois™ Record as depicted in FIGS. **5**E and **5**F. The Whois record, well known and used, contains all of the contact information associated with the person, group, or company that registers a particular domain name. Typically, each Whois is a query and response protocol that is used for querying databases that store an Internet resource's registered users or assignees. It contains information such as the name and contact information of the Registrant (who owns the domain), the name and contact information of the Registrar (the organization or commercial entity that registered the domain name), the registration dates, the name servers, the most recent update, and the expiration date. Whois records may also provide the administrative and technical contact information (which is often, but not always, the registrant). Further, the domain is classified as rogue if authorization is not confirmed based on Whois record. [0106] Identifying the sender associated with rogue domain as malicious and quarantining the message [0107] Communicating a legal notice to the domain owner via Voice user interfaces (VUIs) and over an email, wherein the legal notice is obtained from a legal expert using voice to text technique. [0108] FIG. **5**E-stage **1**-step **1**.**2**-**1** and FIG. **5**F-stage **1**-step **1**.**2**-**1** depict WholsRecordDTD.XML and WholsRecord.XML for ‘example email message type’ respectively. [0109] (ii) However, if the message does not belong to the mass mailer communication and does not have suspected domain name, the body of the message is stripped. Further, as depicted inflow diagram of FIG. **5**H and **51** a checksum is generated from the stripped content of the body to derive presence of mass mailer using checksum hit technique (explained with help of example in Table 2). A heat map of hits is created to check if the hits are beyond a threshold count. If the hits exceed the threshold count, the message (email) is identified as the mass mailer communication and the legal notice is obtained using text to voice technique and communicate to the domain owner using Voice user interfaces (VUIs) and over the email, wherein the sender is identified as potentially malicious.
TABLE-US-00002 TABLE 2 WHOIS hashId hash receiverId hitCounter domain MXRecords Records 2FA rogue AB afc683f51ffa9 AB_receivers.csv 533 domain.co, AB- AB-WHOIS verified TRUE 9a3b3dc0054 domane.com MXR.XML Records.XML 6f54310e4f20 a65e CD bfe6a3f51ffa9 CD_receivers.csv 1 XY.com, CD- CD-WHOIS not- FALSE bc3b3dc0054 XXY.net, MXR.XML Records.XML verified 6f54310e4f20 PXY.org a44f EF 2ce5a3f51ffa9 EF_receivers.csv 52 XZ.com EF- EF-WHOIS verified FALSE 673b3dc0054 MXR.XML Records.XML 6f54310e4f20 a43d [0110] (iii) If the domain IP address is detected to be mobile based on the geo code associated with the sender device present in the set of custom headers and captured in message header, then identify the sender as potentially malicious. NAVIC and/or GPS is used to get lat-long of sender devices (device **104**) used for sending email/messages. The sender device profile helps to collect the specifics of the attackers and the attacker hardware inventory can be profiled accordingly. The domain IP address is identified to be mobile if the geo codes of the sender device when mapped to a physical address and latitude longitude stored in as a time vector varies within a predefined time period such as 1 week, 2 weeks and so on. Explained is the logical reasoning to decide on the time period. The same is depicted in flow diagram of FIG. **5**J It can be understood that when the device is mobile, the device (smart phone) is assigned a IP address via a Dynamic Host Configuration Protocol (DHCP) and does not get a static IP address. DHCP IP addresses have a ‘lease’—that is time validity, after which they expire and hence require renewal or

reissuance. These are automatically handled by the TSP (telecom service provider) in mobile use cases. Also as the devices move between Towers—Cell Locations—there is handover on the IP stack, and IP addresses can change—but migration of data connection takes place.

[0111] Thus, the usable security module of the system **102** utilizes a time vector to detect changes in IP addresses. Thus, DHCP IP if changes indicates the sender device(s) **104** are mobile devices. In certain cases, GPS can be spoofed. If the device profile is known then whether GPS is spoofed can be speculated to a degree. Apps in Indonesia for example spoof device/GPS locations. If a geo field to record the geo codes in the set of custom headers (custom fields) are absent, an API is used to approximate the geo field for the geo codes as depicted in flow diagram of FIG. **5**N. Modifications made by the system **102** (usable security module) to the original email header of FIG. **5**K are shown in FIG. **5**L.

[0112] At step **308** of the method **300**, the one or more hardware processors **204** of the message exchange and communication server **102** (system **102**) by executing the plug-in usable security module are configured by the instructions to process the message using prompt engineering, by the first LLM in conjunction with the phishy-domain-permutations database if the sender of the message is identified as potentially malicious. The phishy-domain-permutations database is pre-generated database as shown in FIG. **4**A (stage **0** of the method **300**) and include steps as explained below:

[0113] Step **1**: Pre-generate, via the second LLM using prompt engineering, permutations of valid non-existent or impersonate domain names in conjunction with each valid domain name and store in the phishy-domain-permutations database. An example

Prompt:

[0114] “Give generic phonetic and visually similar with typeface domain strings representative of Tata.com”

[0115] Response: (for example from LLM such as ChatGPT™) “Phonetically: tata.com Visually: tafa.com”.

[0116] This is done with string matching algorithms which match the permutation string with the senders domain string.

[0117] Examples of phishy-domains generated are provided in Table 3 below and also depicted in FIG. **4**A (Stage **0**—LLM (second LLM)).

TABLE-US-00003 TABLE 3 phishy-domain-permutations.csv permu- domainPermutationId domainId verifiedDomain tation verified p1 d1 tata.com tataa.com FALSE p2 d1 tata.com taata.com FALSE p3 d1 tata.com tara.com TRUE p4 d1 tata.com tala.com TRUE

[0118] Step **2**: Generate a textual and voice communication to the domain owner of each of the valid domain name requesting to register associated permutations of valid non-existent or impersonate domain names and reroute them to valid domain names. The phishy-domain-permutations database is used during the reasoning analysis of LLM to check for presence of invalid domain in the sender domain name.

[0119] The processing of the message at stage **2** of the method **300** (for example the received email as in FIG. **4**B) using prompt engineering, by the first LLM in conjunction with the phishy-domain-permutations database is explained in conjunction with FIGS. **6**A: stage **2**-step**2.1**, **6**B: stage **2**-step **2.2**, **6**C, **6**D and **6**E, stage **2**-step **2.3-1**, **2**,**3-2** and **2.3-3** and FIG. **6**F: stage **2**-step **2.4** respectively. The LLM processing comprises: [0120] (i) Analyzing the message body to provide reasons for sender indicated as malicious, wherein the reasons are in natural language and machine interpretable language. [0121] (ii) Iteratively processing the reasons to generate short reasons excluding technical jargons to be embedded in the body of the message. [0122] (iii) Generating an image using text-to-Image models (for example StableDiffusion), wherein the image is indicative of potentially malicious mail and sentiment in the mail. [0123] (iv) Including a warning indicating IP address is mobile and not fixed.

[0124] At step **310** of the method **300**, the one or more hardware processors **204** of the message

exchange and communication server **102** (system **102**) by executing the plug-in usable security module are configured by the instructions to modify the Document Object Module (DOM) instance for the message to incorporate the processed results of the first LLM. The message is displayed to a receiver with one or more phishing warnings, wherein the phishing warnings are displayed at the DOM nodes based on positions indicated by the first LLM. The positions for the phishing warning indicated by the LLM are in accordance with Usability Principles of Communications that require proximity of warning based on Gestalt principles.). The Usability principle that the usable security module takes into consideration for positioning of the phishing warnings is explained below. User's cognition notices things when they are physically in co-location and ascribe them to the same meaning. For example, a book with a visual on the first page and the text related to it somewhere in the middle of the book. The user will not be able to assign meaning to them as a 'unit'. However, if the same text is immediately below the visual-then the brain treats the text and visual as 'one unit'. It is how the cognitive brain works for a human, and then is applied by the LLM suggestions on positioning of phishing warnings.

[0125] LLM used herein, for example can be any opensource LLMs which are listed on portals such as hugginface.co or github.com or propriety LLMs like Anthropic.

[0126] Stage **3** of the method process corresponds to DOM instance modification to modify email body (message body). Step **3-1** of stage **3** as depicted in FIG. **7**A is example standard Document Object module (DOM) for an email and corresponding email as displayed appears as in FIG. **7**B. The modified DOM is depicted in FIG. **7**C with enclosed dotted lines, with phishing warning message withing enclosed dotted line is also shown. The email from modified DOM as displayed to receiver on device **106** (stage **3**: step **3-3** of the method) appears as depicted in FIG. **7**D. A logical flow for DOM modification is provided in FIG. **7**E. While FIG. **7**F indicates the modified email message as displayed to receiver with warning message inserted.

[0127] The phishing warnings in form of text message are displayed with change the font type face with kerning models as per a device form factor of a receiver device. For example tata.com is not confused with tafa.com based on the typeface. Furthermore, in an embodiment, the message to the receiver with one or more phishing warnings is an audio message if the receiver is registered as visually impaired, and in presented in a local language registered by the receiver if the receiver is a Basic Emergent User (BEU). The methodology to detect whether the user of the device is a BEU is as explained in applicants granted patent application number U.S. Pat. No. 10,620,977B2 titled "Method and system for providing security features in a smart phone." Furthermore, the message displayed to the receiver with the phishing warning comprises a validation functionality (as shown in the UI depicted in FIGS. **8**A through **9**B) enabling the user to trigger the 2FA from the receiver device (device **106**). The user end validation provided allows user (receiver) to establish credibility of the email. The registry has contact [email, phone] of authorized person. In 2FA, the email is updated to say tap/click here for automated 2FA on tapping it using a 'validate' button. An auto callback is generated to the registered number—where in a special key say hash, pound is pressed —this validates the email as authentic on the authorized person side, an automated IVR is setup— which picks up a call and send a hash or pound key automatically—no human is required this email, if required, is auto-authenticated from the domain of the sender organization as is known, this system and number is not with intruder—hence 2FA will fail-so—the message will not be validated. This can be automated, or manual based on the budget of the receiver of the email who will accordingly initiate 2FA.

[0128] This user end additional 2FA validation feature is critical in peer to peer messaging, where the sender and the receiver components are on the same device. In such a network if a malicious user enters, a functionality is provided for the receiver to authenticate the received content. The receiver can use the 'validate' button option on the interface to initiate 2FA.

[0129] This functionality can also be used in client-server mode, for a situation wherein the user assess the need for a manual validation of the received content which may have missed the scrutiny

of the server component containing the usable security module.
[0130] Further, as the system **102** (usable security module) captures data for users clicking malicious links or responding to the potentially malicious messages, this information is recorded to create the personalized threat profile for respective user falling prey to the phishing attack. At step stage **1**-step**1**.**3**-**1** [Table 2], stage **1**-step**1**.**3**-**2** [Table 4] and stage **1**-step**1**.**3**-**3** [Table 5], the one or more hardware processors are configured to create the personalized threat profile at the message exchange and communication server for one or more receivers of the message that are identified as victims if the one or more receivers have read and clicked on one or more messages identified as potentially malicious. The personalized threat profile tags each victim to one or more clusters, and wherein each cluster is associated with word-word phrase having high probability of occurrence (greater than a predefined threshold defined by a Subject Matter Expert) in the potentially malicious message read and clicked by the victim.
[0131] Generation of the personalized threat profile is explained with example using Table 2, Table 4 and Table 5. Table 2 indicates recording of emails that were identified as mass mailer communications and the sender domain was identified as ‘rogue’. Table 4 indicates clustering of receivers into clusters. when ipsum@localhost clicks on the hyper link in the email and falls prey to the phishing message, the system records his action and classifies his into one or more of the clusters. Cluster classification is done using LLM, along with keyword mapping.
[0132] Prompt: From the $body of text identify the emotion that is being expressed and put them into 3 or more clusters of emotions and give label to those clusters.
For example ChatGPT™ response: [0133] Cluster1: Concern [0134] 1. Worry [0135] 2. Distress [0136] Cluster2: Emergency [0137] 1. Urgency [0138] Cluster3: Financial Strain [0139] 1. Shortfall
TABLE-US-00004 TABLE 4 receiverId hash message cluster1 cluster2 clusterN sanjay@tcs.com afc683f51ffa99a3b3dc00546f54310e4f20a65e Hello money money della@tcs.com bfe6a3f51ffa9bc3b3dc00546f54310e4f20a44f Hello lottery lottery sujit@tcs.com 2ce5a3f51ffa9673b3dc00546f54310e4f20a43d Lost money money lost
TABLE-US-00005 TABLE 5 receiverId clusterVulnerability1 clusterVulnerability2 clusterVulnerabilityN sanjay@tcs.com money bank della@tcs.com gift lottery sujit@tcs.com wallet lost
[0140] For example, Sanjay ‘receiver’ is noted to be triggered by words in the message such as money or bank, hence during analysis of potentially malicious mail, when receiver is sanjay@ more focus is provided to mail analysis of receiver sanjay The system **102** is directed to focus via custom prompts engineered and given to the LLM on such emails for Sanjay.
[0141] The personalized threat profile is provided as a feedback during message modification for emphasized protection to the victims for a particular archetype of phishing attack linked with the cluster.
[0142] FIGS. **8**A and **8**B are example interactive User Interface (UIs) with phishing warnings displayed to a receiver on receiver's device by the system for different types of phishing attack via potentially malicious emails detected by the system, in accordance with some embodiments of the present disclosure.
[0143] FIGS. **9**A and **9**B are example interactive User Interface (UIs) with warnings displayed to a receiver on receiver's device by the system for different types of phishing attack via potentially malicious chat messages detected by the system, in accordance with some embodiments of the present disclosure.
[0144] FIGS. **10**A, **10**B, **10**C and **10**D (collectively referred as FIG. **10**) explain stage by stage processing flow by the system for a sample message (instant chat based on Extensible Messaging Presence Protocol-XMPP) received from the sender, wherein the system modifies the sample message to indicate phishing warning before sending it to a receiver, in accordance with some embodiments of the present disclosure. Depicted in the FIG. **10** only those snippets of the chat

message processing which differ from email example depicted in of FIGS. **4**A through **7**F being processed, due specifics of XMPP.

[0145] The chat message over the XMPP protocol is received by an Instant Messaging (IM) server and recorded as in table 6 below.

TABLE-US-00006 TABLE 6 time event from to Jan 04 11:00:10 c2s564b9e449bb0 received lorem@aa$thahealthcare.com ipsum@localhost id message RdLsRMqeg3wDPbn89bf092bda4c26f5723f2d7140bc26a25f625131 <html><body>Our child has met with an accident, and I am in hospital, and I have a shortfall of $$ I am not able to reach your phone please directly pay the hospital at this address. Also, I forgot my phone at home. Sending this from hospital's IM. <ahref='http://aa$thahealthcare.com/ invoce/pay'>http:// aa$thahealthcare.com/invoce/pay make it quick. I am waiting. Regards, Aa$tha Super Specialty Healthcare Majiwada, Thane</body></html>

[0146] FIG. **10**A depicts the message (chat message) received on IM server. As can be seen the XMPP comprises message type stanza used to share chat messages between two parties (sender and receiver). The original message is enclosed with a dotted box marking. It can be noted that like the SPF check in SMTP, IM server performs a Profile Framework (PF) Check, wherein similar to MXRecord Domain Name Server (DNS) Records maintains the legal records with domain name with registrar along with contact details. FIG. **10**B depicts snippet of Snippet of DNRecordDTD.XML and modifications to Chat message header in italics. FIG. **10**C depicts the modified XMPP stanza, highlighted with the dotted box. It can be understood that the process in which the chat message moving from sender to receiver is processed step wise similar to email message processing and repeated for brevity. The differences related to DNS record and stanza of XMPP are depicted in FIG. **10**. FIG. **10**D depicts the modified DOM to include image (depicted with dotted box) indicating phishing warning).

[0147] Thus, the method and system disclosed herein provides intelligent automation of phishing message detection and prevention for all types of users such as BEUs (less literate, basic device users who are not well aware for using the devices and the applications with complete technological understanding), a visually impaired user or any user unable to self-detect a phishing attack via tricky messages by attacker (sender). The method provides the above solution for all IP messaging based protocol message types such as email, Instant Messaging, and the like.

[0148] The written description describes the subject matter herein to enable any person skilled in the art to make and use the embodiments. The scope of the subject matter embodiments is defined by the claims and may include other modifications that occur to those skilled in the art. Such other modifications are intended to be within the scope of the claims if they have similar elements that do not differ from the literal language of the claims or if they include equivalent elements with insubstantial differences from the literal language of the claims.

[0149] It is to be understood that the scope of the protection is extended to such a program and in addition to a computer-readable means having a message therein; such computer-readable storage means contain program-code means for implementation of one or more steps of the method, when the program runs on a server or mobile device or any suitable programmable device. The hardware device can be any kind of device which can be programmed including e.g. any kind of computer like a server or a personal computer, or the like, or any combination thereof. The device may also include means which could be e.g. hardware means like e.g. an application-specific integrated circuit (ASIC), a field-programmable gate array (FPGA), or a combination of hardware and software means, e.g. an ASIC and an FPGA, or at least one microprocessor and at least one memory with software processing components located therein. Thus, the means can include both hardware means, and software means. The method embodiments described herein could be implemented in hardware and software. The device may also include software means. Alternatively, the embodiments may be implemented on different hardware devices, e.g. using a plurality of CPUs.

[0150] The embodiments herein can comprise hardware and software elements. The embodiments that are implemented in software include but are not limited to, firmware, resident software, microcode, etc. The functions performed by various components described herein may be implemented in other components or combinations of other components. For the purposes of this description, a computer-usable or computer readable medium can be any apparatus that can comprise, store, communicate, propagate, or transport the program for use by or in connection with the instruction execution system, apparatus, or device.
[0151] The illustrated steps are set out to explain the exemplary embodiments shown, and it should be anticipated that ongoing technological development will change the manner in which particular functions are performed. These examples are presented herein for purposes of illustration, and not limitation. Further, the boundaries of the functional building blocks have been arbitrarily defined herein for the convenience of the description. Alternative boundaries can be defined so long as the specified functions and relationships thereof are appropriately performed. Alternatives (including equivalents, extensions, variations, deviations, etc., of those described herein) will be apparent to persons skilled in the relevant art(s) based on the teachings contained herein. Such alternatives fall within the scope of the disclosed embodiments. Also, the words “comprising,” “having,” “containing,” and “including,” and other similar forms are intended to be equivalent in meaning and be open ended in that an item or items following any one of these words is not meant to be an exhaustive listing of such item or items or meant to be limited to only the listed item or items. It must also be noted that as used herein and in the appended claims, the singular forms “a,” “an,” and “the” include plural references unless the context clearly dictates otherwise.
[0152] Furthermore, one or more computer-readable storage media may be utilized in implementing embodiments consistent with the present disclosure. A computer-readable storage medium refers to any type of physical memory on which information or data readable by a processor may be stored. Thus, a computer-readable storage medium may store instructions for execution by one or more processors, including instructions for causing the processor(s) to perform steps or stages consistent with the embodiments described herein. The term “computer-readable medium” should be understood to include tangible items and exclude carrier waves and transient signals, i.e., be non-transitory. Examples include random access memory (RAM), read-only memory (ROM), volatile memory, nonvolatile memory, hard drives, CD ROMs, DVDs, flash drives, disks, and any other known physical storage media.
[0153] It is intended that the disclosure and examples be considered as exemplary only, with a true scope of disclosed embodiments being indicated by the following claims.

Claims

1. A processor implemented method for providing usable security, the method comprising: initiating, for each message among a plurality of messages received from one or more senders associated with sender ID and geo-sensor equipped sender devices, a Sender Profile Framework (SPF) check by checking records storing authorized domain names, via a plugin usable security module executed by one or more hardware processors of a message exchange and communication server, wherein the sender ID tagged as SPF-fail indicates that a domain name associated with the sender ID belongs to an unauthorized domain IP address, wherein a SPF stores a legal record of a plurality of domains and associated domain owner and organization details; modifying, by the one or more hardware processors, the records with a set of custom headers comprising a plurality of parameters associated with a sender device profile, wherein the plurality of parameters comprising geocoding, device type, device type profile, sensor type, sensor type profile and accurate real-time positioning and timing services received from satellite agencies; analyzing, by the one or more hardware processors, the one or more sender IDs tagged with SPF-fail in conjunction with one or more of the set of custom headers by performing one of: (i) initiating a two factor authentication

(2FA), if the message is associated with at least one of a mass mailer communication and an unauthorized domain IP address, the 2FA comprises: communicating with a domain owner to confirm authorization of the unauthorized domain IP address, wherein the domain is classified as rogue if authorization is not confirmed, and identifying the sender associated with rogue domain as malicious and quarantining the message; (ii) stripping body of the message, if the message does not belong to the mass mailer communication and does not have suspected domain name, and generating a checksum to derive presence of the mass mailer communication using checksum hit technique, wherein if the hits are beyond a threshold count the sender is identified as potentially malicious; and (iii) if the domain IP address is detected to be mobile based on the geo code associated with the sender device present in the set of custom headers and captured in message header, then identify the sender as potentially malicious; processing, by the one or more hardware processors, the message using prompt engineering, by a first LLM in conjunction with a phishy-domain-permutations database, if the sender of the message is identified as potentially malicious, wherein the processing comprises: (i) analyzing the message body to provide reasons for sender indicated as malicious, wherein the reasons are in natural language and machine interpretable language; (ii) iteratively processing the reasons to generate short reasons excluding technical jargons to be embedded in the body of the message; (iii) generating an image using text-to-Image models, wherein the image is indicative of potentially malicious mail and sentiment in the mail; and (iv) including a warning indicating IP address is mobile and not fixed; and modifying, by the one or more hardware processors, the Document Object Module (DOM) instance for the message to incorporate the processed results of the LLM associated with phishing warnings, wherein the message is displayed to a receiver on a receiver device with one or more phishing warnings, wherein the phishing warnings are displayed at the DOM nodes based on positions indicated by the LLM, and wherein phishing warnings in form of text message are displayed with change the font type face with kerning models as per a device form factor of a receiver device.

2. The method of claim 1, wherein the method comprises creating a personalized threat profile at the message exchange and communication server for one or more receivers of the message that are identified as victims if the one or more receivers have read and clicked on one or more messages identified as potentially malicious, wherein the personalized threat profile tags each victim to one or more clusters, and wherein each cluster is associated with word-word phrase having probability of occurrence greater than a predefined threshold in the potentially malicious message read and clicked by the victim.

3. The method of claim 1, wherein the personalized threat profile is provided as a feedback during message modification for emphasized protection to the victims for a particular archetype of phishing attack linked with the cluster.

4. The method of claim 2, wherein the domain IP address is identified to be mobile if the geo codes of the sender device when mapped to a physical address and latitude longitude stored in as a time vector varies within a predefined time period, wherein if a geo field to record the geo codes in the set of custom headers are absent, an API is used to approximate the geo field for the geo codes.

5. The method of claim 1, wherein communication with the domain owner is triggered, via a voice bot executing a Voice User Interface (VUI), based on information extracted from the legal record in the SPF, wherein a legal notice is communicated to the domain owner via the VUIs and over an email, wherein the legal notice is obtained from a legal expert using voice to text technique, and wherein, in the checksum hit technique, if the hits are beyond a threshold count the legal notice is obtained using text to voice technique and communicated to the domain owner using the VUI and over the email.

6. The method of claim 1, wherein the method comprises: pre-generating via a second LLM permutations of valid non-existent or impersonate domain names in conjunction with each valid domain name and storing in the phishy-domain-permutations database; and generate a textual and voice communication to the domain owner of each of the valid domain name requesting to register

associated permutations of valid non-existent or impersonate domain names and rerouting to valid domain names, wherein the phishy-domain-permutations database is used during a reasoning analysis of LLM to check for presence of invalid domain in the sender domain name.

7. The method of claim 1, wherein the message to the receiver with one or more phishing warnings is an audio message if the receiver is registered as visually impaired, and in presented in a local language registered by the receiver if the receiver is a Basic Emergent User (BEU), wherein positions for the phishing warning indicated by the LLM are in accordance with Usability Principles of Communications that require proximity of warning based on Gestalt principles, and wherein the message displayed to the receiver with the phishing warning comprises a validation functionality enabling the user to trigger the 2FA from the receiver device.

8. A system for providing usable security, the system comprising: a memory storing instructions; one or more Input/Output (I/O) interfaces (**106**); and one or more hardware processors (**104**) coupled to the memory (**102**) via the one or more I/O interfaces (**106**), wherein the one or more hardware processors (**104**) are configured by the instructions to: initiate, for each message among a plurality of messages received from one or more senders associated with sender ID and geo-sensor equipped sender devices, a Sender Profile Framework (SPF) check by checking records storing authorized domain names, via a plugin usable security module executed by a message exchange and communication server, wherein the sender ID tagged as SPF-fail indicates that a domain name associated with the sender ID belongs to an unauthorized domain IP address, wherein a SPF stores legal record of a plurality of domains and associated domain owner and organization details; modify the records with a set of custom headers comprising a plurality of parameters associated with a sender device profile, wherein the plurality of parameters comprising geocoding, device type, device type profile, sensor type, sensor type profile and accurate real-time positioning and timing services received from satellite agencies; analyze, the one or more sender IDs tagged with SPF-fail in conjunction with one or more of the set of custom headers by performing one of: (i) initiating a two factor authentication (2FA), if the message is associated with at least one of a mass mailer communication and an unauthorized domain IP address, the 2FA comprises: communicating with a domain owner to confirm authorization of the unauthorized domain IP address, wherein the domain is classified as rogue if authorization is not confirmed, and identifying the sender associated with rogue domain as malicious and quarantining the message; (ii) stripping body of the message, if the message does not belong to the mass mailer communication and does not have suspected domain name, and generating a checksum to derive presence of the mass mailer communication using checksum hit technique, wherein if the hits are beyond a threshold count the sender is identified as potentially malicious; and (iii) if the domain IP address is detected to be mobile based on the geo code associated with the sender device present in the set of custom headers and captured in message header, then identify the sender as potentially malicious; process the message using prompt engineering, by a first LLM in conjunction with a phishy-domain-permutations database, if the sender of the message is identified as potentially malicious, wherein the processing comprises: (i) analyzing the message body to provide reasons for sender indicated as malicious, wherein the reasons are in natural language and machine interpretable language; (ii) iteratively processing the reasons to generate short reasons excluding technical jargons to be embedded in the body of the message; (iii) generating an image using text-to-Image models, wherein the image is indicative of potentially malicious mail and sentiment in the mail; and (iv) including a warning indicating IP address is mobile and not fixed; and modify the Document Object Module (DOM) instance for the message to incorporate the processed results of the LLM associated with phishing warnings, wherein the message is displayed to a receiver on a receiver device with one or more phishing warnings, wherein the phishing warnings are displayed at the DOM nodes based on positions indicated by the LLM, and wherein phishing warnings in form of text message are displayed with change the font type face with kerning models as per a device form factor of a receiver device.

9. The system of claim 8, wherein the one or more hardware processors are configured to create a personalized threat profile at the message exchange and communication server for one or more receivers of the message that are identified as victims if the one or more receivers have read and clicked on one or more messages identified as potentially malicious, wherein the personalized threat profile tags each victim to one or more clusters, and wherein each cluster is associated with word-word phrase having probability of occurrence greater than a predefined threshold in the potentially malicious message read and clicked by the victim.

10. The system of claim 9, wherein the personalized threat profile is provided as a feedback during message modification for emphasized protection to the victims for a particular archetype of phishing attack linked with the cluster.

11. The system of claim 9, wherein communication with the domain owner is performed, via a voice bot executing a Voice User Interface (VUI), based on information extracted from the legal record in the SPF, wherein a legal notice is communicated to the domain owner via the VUIs and over an email, wherein the legal notice is obtained from a legal expert using voice to text technique, and wherein, in the checksum hit technique, if the hits are beyond a threshold count the legal notice is obtained using text to voice technique and communicated to the domain owner using the VUI and over the email.

12. The system of claim 8, wherein the domain IP address is identified to be mobile if the geo codes of the sender device when mapped to a physical address and latitude longitude stored in as a time vector varies within a predefined time period, wherein if a geo field to record the geo codes in the set of custom headers are absent, an API is used to approximate the geo field for the geo codes.

13. The system of claim 8, wherein the one or more hardware processors are configured to: pre-generate via a second LLM permutations of valid non-existent or impersonate domain names in conjunction with each valid domain name and store in the phishy-domain-permutations database; and generate a textual and voice communication to the domain owner of each of the valid domain name requesting to register associated permutations of valid non-existent or impersonate domain names and reroute to valid domain names, wherein the phishy-domain-permutations database is used during the reasoning analysis of LLM to check for presence of invalid domain in the sender domain name.

14. The system of claim 8, wherein the message to the receiver with one or more phishing warnings is an audio message if the receiver is registered as visually impaired, and in presented in a local language registered by the receiver if the receiver is a Basic Emergent User (BEU), wherein positions for the phishing warning indicated by the LLM are in accordance with Usability Principles of Communications that require proximity of warning based on Gestalt principles, and wherein the message displayed to the receiver with the phishing warning comprises a validation functionality enabling the user to trigger the 2FA from the receiver device.

15. One or more non-transitory machine-readable information storage mediums comprising one or more instructions which when executed by one or more hardware processors cause: initiating, for each message among a plurality of messages received from one or more senders associated with sender ID and geo-sensor equipped sender devices, a Sender Profile Framework (SPF) check by checking records storing authorized domain names, via a plugin usable security module executed by the one or more hardware processors of a message exchange and communication server, wherein the sender ID tagged as SPF-fail indicates that a domain name associated with the sender ID belongs to an unauthorized domain IP address, wherein a SPF stores a legal record of a plurality of domains and associated domain owner and organization details; modifying the records with a set of custom headers comprising a plurality of parameters associated with a sender device profile, wherein the plurality of parameters comprising geocoding, device type, device type profile, sensor type, sensor type profile and accurate real-time positioning and timing services received from satellite agencies; analyzing the one or more sender IDs tagged with SPF-fail in conjunction with one or more of the set of custom headers by performing one of: (i) initiating a two factor

authentication (2FA), if the message is associated with at least one of a mass mailer communication and an unauthorized domain IP address, the 2FA comprises: communicating with a domain owner to confirm authorization of the unauthorized domain IP address, wherein the domain is classified as rogue if authorization is not confirmed, and identifying the sender associated with rogue domain as malicious and quarantining the message; (ii) stripping body of the message, if the message does not belong to the mass mailer communication and does not have suspected domain name, and generating a checksum to derive presence of the mass mailer communication using checksum hit technique, wherein if the hits are beyond a threshold count the sender is identified as potentially malicious; and (iii) if the domain IP address is detected to be mobile based on the geo code associated with the sender device present in the set of custom headers and captured in message header, then identify the sender as potentially malicious; processing, the message using prompt engineering, by a first LLM in conjunction with a phishy-domain-permutations database, if the sender of the message is identified as potentially malicious, wherein the processing comprises: (i) analyzing the message body to provide reasons for sender indicated as malicious, wherein the reasons are in natural language and machine interpretable language; (ii) iteratively processing the reasons to generate short reasons excluding technical jargons to be embedded in the body of the message; (iii) generating an image using text-to-Image models, wherein the image is indicative of potentially malicious mail and sentiment in the mail; and (iv) including a warning indicating IP address is mobile and not fixed; and modifying the Document Object Module (DOM) instance for the message to incorporate the processed results of the LLM associated with phishing warnings, wherein the message is displayed to a receiver on a receiver device with one or more phishing warnings, wherein the phishing warnings are displayed at the DOM nodes based on positions indicated by the LLM, and wherein phishing warnings in form of text message are displayed with change the font type face with kerning models as per a device form factor of a receiver device.

16. The one or more non-transitory machine-readable information storage mediums of claim 15 comprises creating a personalized threat profile at the message exchange and communication server for one or more receivers of the message that are identified as victims if the one or more receivers have read and clicked on one or more messages identified as potentially malicious, wherein the personalized threat profile tags each victim to one or more clusters, and wherein each cluster is associated with word-word phrase having probability of occurrence greater than a predefined threshold in the potentially malicious message read and clicked by the victim.

17. The one or more non-transitory machine-readable information storage mediums of claim 16, wherein the personalized threat profile is provided as a feedback during message modification for emphasized protection to the victims for a particular archetype of phishing attack linked with the cluster.

18. The one or more non-transitory machine-readable information storage mediums of claim 15, wherein the domain IP address is identified to be mobile if the geo codes of the sender device when mapped to a physical address and latitude longitude stored in as a time vector varies within a predefined time period, wherein if a geo field to record the geo codes in the set of custom headers are absent, an API is used to approximate the geo field for the geo codes.

19. The one or more non-transitory machine-readable information storage mediums of claim 15, wherein communication with the domain owner is triggered, via a voice bot executing a Voice User Interface (VUI), based on information extracted from the legal record in the SPF, wherein a legal notice is communicated to the domain owner via the VUIs and over an email, wherein the legal notice is obtained from a legal expert using voice to text technique, and wherein, in the checksum hit technique, if the hits are beyond a threshold count the legal notice is obtained using text to voice technique and communicated to the domain owner using the VUI and over the email.

20. The one or more non-transitory machine-readable information storage mediums claim 15 comprising: pre-generating via a second LLM permutations of valid non-existent or impersonate domain names in conjunction with each valid domain name and storing in the phishy-domain-

permutations database; and generate a textual and voice communication to the domain owner of each of the valid domain name requesting to register associated permutations of valid non-existent or impersonate domain names and rerouting to valid domain names, wherein the phishy-domain-permutations database is used during a reasoning analysis of LLM to check for presence of invalid domain in the sender domain name.