SYSTEM AND METHOD FOR SEARCHING AND RETRIEVING CERTIFICATES

[0001] The invention relates generally to the processing of messages, such as e-mail messages, and more specifically to a system and method for searching and retrieving certificates used in the processing of encoded messages.

Background of the Invention

[0002] Electronic mail ("e-mail") messages may be encoded using one of a number of known protocols. Some of these protocols, such as Secure Multiple Internet Mail Extensions ("S/MIME") for example, rely on public and private encryption keys to provide confidentiality and integrity, and on a Public Key Infrastructure (PKI) to communicate information that provides authentication and authorization. Data encrypted using a private key of a private key/public key pair can only be decrypted using the corresponding public key of the pair, and vice-versa. The authenticity of public keys used in the encoding of messages is validated using certificates. In particular, if a user of a computing device wishes to encrypt a message before the message is sent to a particular individual, the user will require a certificate for that individual. That certificate will typically comprise the public key of the individual, as well as other identification-related information.

[0003] If the requisite certificate for the intended recipient is not already stored on the user's computing device, the certificate must first be retrieved. Searching for and retrieving a certificate for a specific recipient is a process that generally involves querying a certificate server by having the user manually enter the name and/or e-mail address of the intended recipient in a search form displayed on the computing device. Generally, certificates located in the search are then temporarily downloaded to the computing device for consideration, and a list of located certificates may then be displayed to the user. Selected certificates in the list may then be manually identified by a user for storage in a non-volatile store of the computing device, for potential future use.

[0004] In some implementations, however, instead of temporarily downloading all the certificates located in the search to the computing device in the first instance, only certain data needed to generate a list of certificates located in the search may be initially downloaded to the computing device. The list is displayed to a user, typically identifying

each located certificate using the common name and e-mail address of the individual to whom the respective certificate has been issued. Only after the user selects specific certificates from the list to be stored for future use are the certificates downloaded to the computing device for storage. In particular, if the computing device is a mobile device, deferring the downloading of certificates to the mobile device, and only downloading the user-selected certificates, can significantly minimize waste of resources.

[0005] Unfortunately, in these implementations where the downloading of certificates is deferred, it is generally not possible to determine with certainty whether any of the certificates identified in the list are already stored in a certificate store of the computing device from only the downloaded data used to generate the list. For example, in order for an application on the computing device to be able to determine with certainty that a given certificate identified in the list is already stored in the certificate store, the actual certificate would typically need to be downloaded to the computing device so that the requisite data needed to make the determination can be obtained. This can be a time-consuming and expensive task (e.g. with respect to bandwidth), and may be wasteful if the downloaded certificate is, in fact, already in the certificate store.

Summary of the Invention

[0006] Embodiments of the invention are generally directed to a system and method for more efficiently searching certificates on a computing device and retrieving certificates for storage on the computing device.

[0007] In one broad aspect of the invention, there is provided a method of searching and retrieving certificates comprising the steps of: receiving a certificate search request; performing a search on one or more certificate servers, wherein at least one query is submitted to the one or more certificate servers to request retrieval of certificates satisfying the certificate search request; retrieving at least one certificate from the one or more certificate servers; processing each certificate retrieved at a first computing device to determine search result data, wherein the search result data comprises data that uniquely identifies the respective certificate; and for each certificate retrieved, communicating search result data associated therewith to a second device from the first computing device, for use in determining whether the respective certificate is stored on the second device.

[0008] In another broad aspect of the invention, the data that uniquely identifies each respective certificate comprises serial number and issuer data for the respective certificate, and the processing step comprises parsing each certificate retrieved to obtain the respective serial number and issuer data.

[0009] In another broad aspect of the invention, the data that uniquely identifies each respective certificate comprises a hash of at least a part of the respective certificate, and the processing step comprises applying a hash algorithm to each certificate retrieved to obtain the respective hash.

[0010] In another broad aspect of the invention, there is provided a system for searching and retrieving certificates comprising at least a first computing device and a second device, wherein the first computing device is adapted to: receive a certificate search request; perform a search on one or more certificate servers by submitting at least one query thereto to request retrieval of certificates satisfying the certificate search request; retrieve at least one certificate from the one or more certificate servers; process each certificate retrieved to determine search result data, wherein the search result data comprises data that uniquely identifies the respective certificate; and for each certificate retrieved, communicate search result data associated therewith to a second device for use in determining whether the respective certificate is stored on the second device.

Brief Description of the Drawings

[0011] For a better understanding of embodiments of the invention, and to show more clearly how it may be carried into effect, reference will now be made, by way of example, to the accompanying drawings in which:

FIG. 1 is a block diagram of a mobile device in one example implementation;

FIG. 2 is a block diagram of a communication subsystem component of the mobile device of FIG. 1;

FIG. 3 is a block diagram of a node of a wireless network;

FIG. 4 is a block diagram illustrating components of a host system in one example configuration;

FIG. 5 is a block diagram showing an example of a certificate chain;

FIG. 6 is a block diagram illustrating components of an example of an encoded message;

FIG. 7A is a flowchart illustrating steps in a method of searching and retrieving certificates in an embodiment of the invention; and

FIG. 7B is a flowchart illustrating steps in a method of searching and retrieving certificates in another embodiment of the invention.

Detailed Description of Embodiments of the Invention

[0012] Some embodiments of the invention make use of a mobile station. A mobile station is a two-way communication device with advanced data communication capabilities having the capability to communicate with other computer systems, and is also referred to herein generally as a mobile device. A mobile device may also include the capability for voice communications. Depending on the functionality provided by a mobile device, it may be referred to as a data messaging device, a two-way pager, a cellular telephone with data messaging capabilities, a wireless Internet appliance, or a data communication device (with or without telephony capabilities). A mobile device communicates with other devices through a network of transceiver stations.

[0013] To aid the reader in understanding the structure of a mobile device and how it communicates with other devices, reference is made to FIGS. 1 through 3.

[0014] Referring first to FIG. 1, a block diagram of a mobile device in one example implementation is shown generally as 100. Mobile device 100 comprises a number of components, the controlling component being microprocessor 102. Microprocessor 102 controls the overall operation of mobile device 100. Communication functions, including data and voice communications, are performed through communication subsystem 104. Communication subsystem 104 receives messages from and sends messages to a wireless network 200. In this example implementation of mobile device 100, communication subsystem 104 is configured in accordance with the Global System for Mobile Communication (GSM) and General Packet Radio Services (GPRS) standards. The GSM/GPRS wireless network is used worldwide and it is expected that these standards will be superseded eventually by Enhanced Data GSM Environment (EDGE) and Universal Mobile Telecommunications Service (UMTS). New standards are still being defined, but it is believed that they will have similarities to the network

behaviour described herein, and it will also be understood by persons skilled in the art that the invention is intended to use any other suitable standards that are developed in the future. The wireless link connecting communication subsystem 104 with network 200 represents one or more different Radio Frequency (RF) channels, operating according to defined protocols specified for GSM/GPRS communications. With newer network protocols, these channels are capable of supporting both circuit switched voice communications and packet switched data communications.

[0015] Although the wireless network associated with mobile device 100 is a GSM/GPRS wireless network in one example implementation of mobile device 100, other wireless networks may also be associated with mobile device 100 in variant implementations. Different types of wireless networks that may be employed include, for example, data-centric wireless networks, voice-centric wireless networks, and dual-mode networks that can support both voice and data communications over the same physical base stations. Combined dual-mode networks include, but are not limited to, Code Division Multiple Access (CDMA) or CDMA2000 networks, GSM/GPRS networks (as mentioned above), and future third-generation (3G) networks like EDGE and UMTS. Some older examples of data-centric networks include the Mobitex™ Radio Network and the DataTAC™ Radio Network. Examples of older voice-centric data networks include Personal Communication Systems (PCS) networks like GSM and Time Division Multiple Access (TDMA) systems.

[0016] Microprocessor 102 also interacts with additional subsystems such as a Random Access Memory (RAM) 106, flash memory 108, display 110, auxiliary input/output (I/O) subsystem 112, serial port 114, keyboard 116, speaker 118, microphone 120, short-range communications 122 and other devices 124.

[0017] Some of the subsystems of mobile device 100 perform communication-related functions, whereas other subsystems may provide "resident" or on-device functions. By way of example, display 110 and keyboard 116 may be used for both communication-related functions, such as entering a text message for transmission over network 200, and device-resident functions such as a calculator or task list. Operating system software used by microprocessor 102 is typically stored in a persistent store such as flash memory 108, which may alternatively be a read-only memory (ROM) or similar storage element (not shown). Those skilled in the art will appreciate that the operating

system, specific device applications, or parts thereof, may be temporarily loaded into a volatile store such as RAM 106.

[0018] Mobile device 100 may send and receive communication signals over network 200 after required network registration or activation procedures have been completed. Network access is associated with a subscriber or user of a mobile device 100. To identify a subscriber, mobile device 100 requires a Subscriber Identity Module or "SIM" card 126 to be inserted in a SIM interface 128 in order to communicate with a network. SIM 126 is one type of a conventional "smart card" used to identify a subscriber of mobile device 100 and to personalize the mobile device 100, among other things. Without SIM 126, mobile device 100 is not fully operational for communication with network 200. By inserting SIM 126 into SIM interface 128, a subscriber can access all subscribed services. Services could include: web browsing and messaging such as e-mail, voice mail, Short Message Service (SMS), and Multimedia Messaging Services (MMS). More advanced services may include: point of sale, field service and sales force automation. SIM 126 includes a processor and memory for storing information. Once SIM 126 is inserted in SIM interface 128, it is coupled to microprocessor 102. In order to identify the subscriber, SIM 126 contains some user parameters such as an International Mobile Subscriber Identity (IMSI). An advantage of using SIM 126 is that a subscriber is not necessarily bound by any single physical mobile device. SIM 126 may store additional subscriber information for a mobile device as well, including datebook (or calendar) information and recent call information.

[0019] Mobile device 100 is a battery-powered device and includes a battery interface 132 for receiving one or more rechargeable batteries 130. Battery interface 132 is coupled to a regulator (not shown), which assists battery 130 in providing power V+ to mobile device 100. Although current technology makes use of a battery, future technologies such as micro fuel cells may provide the power to mobile device 100.

[0020] Microprocessor 102, in addition to its operating system functions, enables execution of software applications on mobile device 100. A set of applications that control basic device operations, including data and voice communication applications, will normally be installed on mobile device 100 during its manufacture. Another application that may be loaded onto mobile device 100 would be a personal information manager (PIM). A PIM has functionality to organize and manage data items of interest to a

subscriber, such as, but not limited to, e-mail, calendar events, voice mails, appointments, and task items. A PIM application has the ability to send and receive data items via wireless network 200. PIM data items may be seamlessly integrated, synchronized, and updated via wireless network 200 with the mobile device subscriber's corresponding data items stored and/or associated with a host computer system. This functionality creates a mirrored host computer on mobile device 100 with respect to such items. This can be particularly advantageous where the host computer system is the mobile device subscriber's office computer system.

[0021] Additional applications may also be loaded onto mobile device 100 through network 200, auxiliary I/O subsystem 112, serial port 114, short-range communications subsystem 122, or any other suitable subsystem 124. This flexibility in application installation increases the functionality of mobile device 100 and may provide enhanced on-device functions, communication-related functions, or both. For example, secure communication applications may enable electronic commerce functions and other such financial transactions to be performed using mobile device 100.

[0022] Serial port 114 enables a subscriber to set preferences through an external device or software application and extends the capabilities of mobile device 100 by providing for information or software downloads to mobile device 100 other than through a wireless communication network. The alternate download path may, for example, be used to load an encryption key onto mobile device 100 through a direct and thus reliable and trusted connection to provide secure device communication.

[0023] Short-range communications subsystem 122 provides for communication between mobile device 100 and different systems or devices, without the use of network 200. For example, subsystem 122 may include an infrared device and associated circuits and components for short-range communication. Examples of short range communication would include standards developed by the Infrared Data Association (IrDA), Bluetooth, and the 802.11 family of standards developed by IEEE.

[0024] In use, a received signal such as a text message, an e-mail message, or web page download will be processed by communication subsystem 104 and input to microprocessor 102. Microprocessor 102 will then process the received signal for output to display 110 or alternatively to auxiliary I/O subsystem 112. A subscriber may also compose data items, such as e-mail messages, for example, using keyboard 116 in

conjunction with display 110 and possibly auxiliary I/O subsystem 112. Auxiliary subsystem 112 may include devices such as: a touch screen, mouse, track ball, infrared fingerprint detector, or a roller wheel with dynamic button pressing capability. Keyboard 116 is an alphanumeric keyboard and/or telephone-type keypad. A composed item may be transmitted over network 200 through communication subsystem 104.

[0025] For voice communications, the overall operation of mobile device 100 is substantially similar, except that the received signals would be output to speaker 118, and signals for transmission would be generated by microphone 120. Alternative voice or audio I/O subsystems, such as a voice message recording subsystem, may also be implemented on mobile device 100. Although voice or audio signal output is accomplished primarily through speaker 118, display 110 may also be used to provide additional information such as the identity of a calling party, duration of a voice call, or other voice call related information.

[0026] Referring now to FIG. 2, a block diagram of the communication subsystem component 104 of FIG. 1 is shown. Communication subsystem 104 comprises a receiver 150, a transmitter 152, one or more embedded or internal antenna elements 154, 156, Local Oscillators (LOs) 158, and a processing module such as a Digital Signal Processor (DSP) 160.

[0027] The particular design of communication subsystem 104 is dependent upon the network 200 in which mobile device 100 is intended to operate, thus it should be understood that the design illustrated in FIG. 2 serves only as one example. Signals received by antenna 154 through network 200 are input to receiver 150, which may perform such common receiver functions as signal amplification, frequency down conversion, filtering, channel selection, and analog-to-digital (A/D) conversion. A/D conversion of a received signal allows more complex communication functions such as demodulation and decoding to be performed in DSP 160. In a similar manner, signals to be transmitted are processed, including modulation and encoding, by DSP 160. These DSP-processed signals are input to transmitter 152 for digital-to-analog (D/A) conversion, frequency up conversion, filtering, amplification and transmission over network 200 via antenna 156. DSP 160 not only processes communication signals, but also provides for receiver and transmitter control. For example, the gains applied to communication signals

in receiver 150 and transmitter 152 may be adaptively controlled through automatic gain control algorithms implemented in DSP 160.

[0028] The wireless link between mobile device 100 and a network 200 may contain one or more different channels, typically different RF channels, and associated protocols used between mobile device 100 and network 200. A RF channel is a limited resource that must be conserved, typically due to limits in overall bandwidth and limited battery power of mobile device 100.

[0029] When mobile device 100 is fully operational, transmitter 152 is typically keyed or turned on only when it is sending to network 200 and is otherwise turned off to conserve resources. Similarly, receiver 150 is periodically turned off to conserve power until it is needed to receive signals or information (if at all) during designated time periods.

[0030] Referring now to FIG. 3, a block diagram of a node of a wireless network is shown as 202. In practice, network 200 comprises one or more nodes 202. Mobile device 100 communicates with a node 202 within wireless network 200. In the example implementation of FIG. 3, node 202 is configured in accordance with General Packet Radio Service (GPRS) and Global Systems for Mobile (GSM) technologies. Node 202 includes a base station controller (BSC) 204 with an associated tower station 206, a Packet Control Unit (PCU) 208 added for GPRS support in GSM, a Mobile Switching Center (MSC) 210, a Home Location Register (HLR) 212, a Visitor Location Registry (VLR) 214, a Serving GPRS Support Node (SGSN) 216, a Gateway GPRS Support Node (GGSN) 218, and a Dynamic Host Configuration Protocol (DHCP) 220. This list of components is not meant to be an exhaustive list of the components of every node 202 within a GSM/GPRS network, but rather a list of components that are commonly used in communications through network 200.

[0031] In a GSM network, MSC 210 is coupled to BSC 204 and to a landline network, such as a Public Switched Telephone Network (PSTN) 222 to satisfy circuit switched requirements. The connection through PCU 208, SGSN 216 and GGSN 218 to the public or private network (Internet) 224 (also referred to herein generally as a shared network infrastructure) represents the data path for GPRS capable mobile devices. In a GSM network extended with GPRS capabilities, BSC 204 also contains a Packet Control Unit (PCU) 208 that connects to SGSN 216 to control segmentation, radio channel

allocation and to satisfy packet switched requirements. To track mobile device location and availability for both circuit switched and packet switched management, HLR 212 is shared between MSC 210 and SGSN 216. Access to VLR 214 is controlled by MSC 210.

[0032] Station 206 is a fixed transceiver station. Station 206 and BSC 204 together form the fixed transceiver equipment. The fixed transceiver equipment provides wireless network coverage for a particular coverage area commonly referred to as a "cell". The fixed transceiver equipment transmits communication signals to and receives communication signals from mobile devices within its cell via station 206. The fixed transceiver equipment normally performs such functions as modulation and possibly encoding and/or encryption of signals to be transmitted to the mobile device in accordance with particular, usually predetermined, communication protocols and parameters, under control of its controller. The fixed transceiver equipment similarly demodulates and possibly decodes and decrypts, if necessary, any communication signals received from mobile device 100 within its cell. Communication protocols and parameters may vary between different nodes. For example, one node may employ a different modulation scheme and operate at different frequencies than other nodes.

[0033] For all mobile devices 100 registered with a specific network, permanent configuration data such as a user profile is stored in HLR 212. HLR 212 also contains location information for each registered mobile device and can be queried to determine the current location of a mobile device. MSC 210 is responsible for a group of location areas and stores the data of the mobile devices currently in its area of responsibility in VLR 214. Further VLR 214 also contains information on mobile devices that are visiting other networks. The information in VLR 214 includes part of the permanent mobile device data transmitted from HLR 212 to VLR 214 for faster access. By moving additional information from a remote HLR 212 node to VLR 214, the amount of traffic between these nodes can be reduced so that voice and data services can be provided with faster response times and at the same time requiring less use of computing resources.

[0034] SGSN 216 and GGSN 218 are elements added for GPRS support; namely packet switched data support, within GSM. SGSN 216 and MSC 210 have similar responsibilities within wireless network 200 by keeping track of the location of each mobile device 100. SGSN 216 also performs security functions and access control for data traffic on network 200. GGSN 218 provides internetworking connections with

external packet switched networks and connects to one or more SGSN's 216 via an Internet Protocol (IP) backbone network operated within the network 200. During normal operations, a given mobile device 100 must perform a "GPRS Attach" to acquire an IP address and to access data services. This requirement is not present in circuit switched voice channels as Integrated Services Digital Network (ISDN) addresses are used for routing incoming and outgoing calls. Currently, all GPRS capable networks use private, dynamically assigned IP addresses, thus requiring a DHCP server 220 connected to the GGSN 218. There are many mechanisms for dynamic IP assignment, including using a combination of a Remote Authentication Dial-In User Service (RADIUS) server and DHCP server. Once the GPRS Attach is complete, a logical connection is established from a mobile device 100, through PCU 208, and SGSN 216 to an Access Point Node (APN) within GGSN 218. The APN represents a logical end of an IP tunnel that can either access direct Internet compatible services or private network connections. The APN also represents a security mechanism for network 200, insofar as each mobile device 100 must be assigned to one or more APNs and mobile devices 100 cannot exchange data without first performing a GPRS Attach to an APN that it has been authorized to use. The APN may be considered to be similar to an Internet domain name such as "myconnection.wireless.com".

[0035] Once the GPRS Attach is complete, a tunnel is created and all traffic is exchanged within standard IP packets using any protocol that can be supported in IP packets. This includes tunneling methods such as IP over IP as in the case with some IPSecurity (IPsec) connections used with Virtual Private Networks (VPN). These tunnels are also referred to as Packet Data Protocol (PDP) Contexts and there are a limited number of these available in the network 200. To maximize use of the PDP Contexts, network 200 will run an idle timer for each PDP Context to determine if there is a lack of activity. When a mobile device 100 is not using its PDP Context, the PDP Context can be deallocated and the IP address returned to the IP address pool managed by DHCP server 220.

[0036] Referring now to FIG. 4, a block diagram illustrating components of a host system in one example configuration is shown. Host system 250 will typically be a corporate office or other local area network (LAN), but may instead be a home office computer or some other private system, for example, in variant implementations. In this

example shown in FIG. 4, host system 250 is depicted as a LAN of an organization to which a user of mobile device 100 belongs.

[0037] LAN 250 comprises a number of network components connected to each other by LAN connections 260. For instance, a user's desktop computer 262a with an accompanying cradle 264 for the user's mobile device 100 is situated on LAN 250. Cradle 264 for mobile device 100 may be coupled to computer 262a by a serial or a Universal Serial Bus (USB) connection, for example. Other user computers 262b are also situated on LAN 250, and each may or may not be equipped with an accompanying cradle 264 for a mobile device. Cradle 264 facilitates the loading of information (e.g. PIM data, private symmetric encryption keys to facilitate secure communications between mobile device 100 and LAN 250) from user computer 262a to mobile device 100, and may be particularly useful for bulk information updates often performed in initializing mobile device 100 for use. The information downloaded to mobile device 100 may include certificates used in the exchange of messages. It will be understood by persons skilled in the art that user computers 262a, 262b will typically be also connected to other peripheral devices not explicitly shown in FIG. 4.

[0038] Embodiments of the invention relate generally to the processing of messages, such as e-mail messages, and some embodiments relate generally to the communication of such messages to and from mobile device 100. Accordingly, only a subset of network components of LAN 250 are shown in FIG. 4 for ease of exposition, and it will be understood by persons skilled in the art that LAN 250 will comprise additional components not explicitly shown in FIG. 4, for this example configuration. More generally, LAN 250 may represent a smaller part of a larger network [not shown] of the organization, and may comprise different components and/or be arranged in different topologies than that shown in the example of FIG. 4.

[0039] In this example, mobile device 100 communicates with LAN 250 through a node 202 of wireless network 200 and a shared network infrastructure 224 such as a service provider network or the public Internet. Access to LAN 250 may be provided through one or more routers [not shown], and computing devices of LAN 250 may operate from behind a firewall or proxy server 266.

[0040] In a variant implementation, LAN 250 comprises a wireless VPN router [not shown] to facilitate data exchange between the LAN 250 and mobile device 100. The

concept of a wireless VPN router is new in the wireless industry and implies that a VPN connection can be established directly through a specific wireless network to mobile device 100. The possibility of using a wireless VPN router has only recently been available and could be used when the new Internet Protocol (IP) Version 6 (IPV6) arrives into IP-based wireless networks. This new protocol will provide enough IP addresses to dedicate an IP address to every mobile device, making it possible to push information to a mobile device at any time. An advantage of using a wireless VPN router is that it could be an off-the-shelf VPN component, not requiring a separate wireless gateway and separate wireless infrastructure to be used. A VPN connection would preferably be a Transmission Control Protocol (TCP)/IP or User Datagram Protocol (UDP)/IP connection to deliver the messages directly to mobile device 100 in this variant implementation.

[0041] Messages intended for a user of mobile device 100 are initially received by a message server 268 of LAN 250. Such messages may originate from any of a number of sources. For instance, a message may have been sent by a sender from a computer 262b within LAN 250, from a different mobile device [not shown] connected to wireless network 200 or to a different wireless network, or from a different computing device or other device capable of sending messages, via the shared network infrastructure 224, and possibly through an application service provider (ASP) or Internet service provider (ISP), for example.

[0042] Message server 268 typically acts as the primary interface for the exchange of messages, particularly e-mail messages, within the organization and over the shared network infrastructure 224. Each user in the organization that has been set up to send and receive messages is typically associated with a user account managed by message server 268. One example of a message server 268 is a Microsoft Exchange™ Server. In some implementations, LAN 250 may comprise multiple message servers 268. Message server 268 may also be adapted to provide additional functions beyond message management, including the management of data associated with calendars and task lists, for example.

[0043] When messages are received by message server 268, they are typically stored in a message store [not explicitly shown], from which messages can be subsequently retrieved and delivered to users. For instance, an e-mail client application operating on a user's computer 262a may request the e-mail messages associated with that

user's account stored on message server 268. These messages would then typically be retrieved from message server 268 and stored locally on computer 262a.

[0044] When operating mobile device 100, the user may wish to have e-mail messages retrieved for delivery to the handheld. An e-mail client application operating on mobile device 100 may also request messages associated with the user's account from message server 268. The e-mail client may be configured (either by the user or by an administrator, possibly in accordance with an organization's information technology (IT) policy) to make this request at the direction of the user, at some pre-defined time interval, or upon the occurrence of some pre-defined event. In some implementations, mobile device 100 is assigned its own e-mail address, and messages addressed specifically to mobile device 100 are automatically redirected to mobile device 100 as they are received by message server 268.

[0045] To facilitate the wireless communication of messages and message-related data between mobile device 100 and components of LAN 250, a number of wireless communications support components 270 may be provided. In this example implementation, wireless communications support components 270 comprise a message management server 272, for example. Message management server 272 is used to specifically provide support for the management of messages, such as e-mail messages, that are to be handled by mobile devices. Generally, while messages are still stored on message server 268, message management server 272 can be used to control when, if, and how messages should be sent to mobile device 100. Message management server 272 also facilitates the handling of messages composed on mobile device 100, which are sent to message server 268 for subsequent delivery.

[0046] For example, message management server 272 may: monitor the user's "mailbox" (e.g. the message store associated with the user's account on message server 268) for new e-mail messages; apply user-definable filters to new messages to determine if and how the messages will be relayed to the user's mobile device 100; compress and encrypt new messages (e.g. using an encryption technique such as Data Encryption Standard (DES) or Triple DES) and push them to mobile device 100 via the shared network infrastructure 224 and wireless network 200; and receive messages composed on mobile device 100 (e.g. encrypted using Triple DES), decrypt and decompress the composed messages, re-format the composed messages if desired so that they will appear

to have originated from the user's computer 262a, and re-route the composed messages to message server 268 for delivery.

[0047] Certain properties or restrictions associated with messages that are to be sent from and/or received by mobile device 100 can be defined (e.g. by an administrator in accordance with IT policy) and enforced by message management server 272. These may include whether mobile device 100 may receive encrypted and/or signed messages, minimum encryption key sizes, whether outgoing messages must be encrypted and/or signed, and whether copies of all secure messages sent from mobile device 100 are to be sent to a pre-defined copy address, for example.

[0048] Message management server 272 may also be adapted to provide other control functions, such as only pushing certain message information or pre-defined portions (e.g. "blocks") of a message stored on message server 268 to mobile device 100. For example, when a message is initially retrieved by mobile device 100 from message server 268, message management server 272 is adapted to push only the first part of a message to mobile device 100, with the part being of a pre-defined size (e.g. 2 KB). The user can then request more of the message, to be delivered in similar-sized blocks by message management server 272 to mobile device 100, possibly up to a maximum pre-defined message size.

[0049] Accordingly, message management server 272 facilitates better control over the type of data and the amount of data that is communicated to mobile device 100, and can help to minimize potential waste of bandwidth or other resources.

[0050] It will be understood by persons skilled in the art that message management server 272 need not be implemented on a separate physical server in LAN 250 or other network. For example, some or all of the functions associated with message management server 272 may be integrated with message server 268, or some other server in LAN 250. Furthermore, LAN 250 may comprise multiple message management servers 272, particularly in variant implementations where a large number of mobile devices need to be supported.

[0051] Embodiments of the invention relate generally to the processing of encoded messages, such as e-mail messages that are encrypted and/or signed. While Simple Mail Transfer Protocol (SMTP), RFC822 headers, and Multipurpose Internet Mail Extensions

(MIME) body parts may be used to define the format of a typical e-mail message not requiring encoding, Secure/MIME (S/MIME), a version of the MIME protocol, may be used in the communication of encoded messages (i.e. in secure messaging applications). S/MIME enables end-to-end authentication and confidentiality, and protects data integrity and privacy from the time an originator of a message sends a message until it is decoded and read by the message recipient. Other known standards and protocols may be employed to facilitate secure message communication, such as Pretty Good Privacy™ (PGP), OpenPGP, and others known in the art.

[0052] Secure messaging protocols such as S/MIME rely on public and private encryption keys to provide confidentiality and integrity, and on a Public Key Infrastructure (PKI) to communicate information that provides authentication and authorization. Data encrypted using a private key of a private key/public key pair can only be decrypted using the corresponding public key of the pair, and vice-versa. Private key information is never made public, whereas public key information is shared.

[0053] For example, if a sender wishes to send a message to a recipient in encrypted form, the recipient's public key is used to encrypt a message, which can then be decrypted only using the recipient's private key. Alternatively, in some encoding techniques, a one-time session key is generated and used to encrypt the body of a message, typically with a symmetric encryption technique (e.g. Triple DES). The session key is then encrypted using the recipient's public key (e.g. with a public key encryption algorithm such as RSA), which can then be decrypted only using the recipient's private key. The decrypted session key can then be used to decrypt the message body. The message header may be used to specify the particular encryption scheme that must be used to decrypt the message. Other encryption techniques based on public key cryptography may be used in variant implementations. However, in each of these cases, only the recipient's private key may be used to facilitate decryption of the message, and in this way, the confidentiality of messages can be maintained.

[0054] As a further example, a sender may sign a message using a digital signature. A digital signature is a digest of the message (e.g. a hash of the message) encrypted using the sender's private key, which can then be appended to the outgoing message. To verify the signature of the message when received, the recipient uses the same technique as the sender (e.g. using the same standard hash algorithm) to obtain a

digest of the received message. The recipient also uses the sender's public key to decrypt the digital signature, in order to obtain what should be a matching digest for the received message. If the digests of the received message do not match, this suggests that either the message content was changed during transport and/or the message did not originate from the sender whose public key was used for verification. By verifying a digital signature in this way, authentication of the sender and message integrity can be maintained.

[0055] An encoded message may be encrypted, signed, or both encrypted and signed. The authenticity of public keys used in these operations is validated using certificates. A certificate is a digital document issued by a certificate authority (CA). Certificates are used to authenticate the association between users and their public keys, and essentially, provides a level of trust in the authenticity of the users' public keys. Certificates contain information about the certificate holder, with certificate contents typically formatted in accordance with an accepted standard (e.g. X.509).

[0056] Consider FIG. 5, in which an example certificate chain 300 is shown. Certificate 310 issued to "John Smith" is an example of a certificate issued to an individual, which may be referred to as an end entity certificate. End entity certificate 310 typically identifies the certificate holder 312 (i.e. John Smith in this example) and the issuer of the certificate 314, and includes a digital signature of the issuer 316 and the certificate holder's public key 318. Certificate 310 will also typically include other information and attributes that identify the certificate holder (e.g. e-mail address, organization name, organizational unit name, location, etc.). When the individual composes a message to be sent to a recipient, it is customary to include that individual's certificate 300 with the message.

[0057] For a public key to be trusted, its issuing organization must be trusted. The relationship between a trusted CA and a user's public key can be represented by a series of related certificates, also referred to as a certificate chain. The certificate chain can be followed to determine the validity of a certificate.

[0058] For instance, in the example certificate chain 300 shown in FIG. 5, the recipient of a message purported to be sent by John Smith may wish to verify the trust status of certificate 310 attached to the received message. To verify the trust status of certificate 310 on a recipient's computing device (e.g. computer 262a of FIG. 4) for example, the certificate 320 of issuer ABC is obtained, and used to verify that certificate

310 was indeed signed by issuer ABC. Certificate 320 may already be stored in a certificate store on the computing device, or it may need to be retrieved from a certificate source (e.g. LDAP server 284 of FIG. 4 or some other public or private LDAP server). If certificate 320 is already stored in the recipient's computing device and the certificate has been designated as trusted by the recipient, then certificate 310 is considered to be trusted since it chains to a stored, trusted certificate.

[0059] However, in the example shown in FIG. 5, certificate 330 is also required to verify the trust of certificate 310. Certificate 330 is self-signed, and is referred to as a "root certificate". Accordingly, certificate 320 may be referred to as an "intermediate certificate" in certificate chain 300; any given certificate chain to a root certificate, assuming a chain to the root certificate can be determined for a particular end entity certificate, may contain zero, one, or multiple intermediate certificates. If certificate 330 is a root certificate issued by a trusted source (from a large certificate authority such as Verisign or Entrust, for example), then certificate 310 may be considered to be trusted since it chains to a trusted certificate. The implication is that both the sender and the recipient of the message trust the source of the root certificate 330. If a certificate cannot be chained to a trusted certificate, the certificate may be considered to be "not trusted".

[0060] Certificate servers store information about certificates and lists identifying certificates that have been revoked. These certificate servers can be accessed to obtain certificates and to verify certificate authenticity and revocation status. For example, a Lightweight Directory Access Protocol (LDAP) server may be used to obtain certificates, and an Online Certificate Status Protocol (OCSP) server may be used to verify certificate revocation status.

[0061] Standard e-mail security protocols typically facilitate secure message transmission between non-mobile computing devices (e.g. computers 262a, 262b of FIG. 4; remote desktop devices). Referring again to FIG. 4, in order that signed messages received from senders may be read from mobile device 100 and encrypted messages be sent to those senders, mobile device 100 is adapted to store certificates and associated public keys of other individuals. Certificates stored on a user's computer 262a will typically be downloaded from computer 262a to mobile device 100 through cradle 264, for example.

[0062] Certificates stored on computer 262a and downloaded to mobile device 100 are not limited to certificates associated with individuals but may also include certificates issued to CAs, for example. Certain certificates stored in computer 262a and/or mobile device 100 can also be explicitly designated as "trusted" by the user. Accordingly, when a certificate is received by a user on mobile device 100, it can be verified on mobile device 100 by matching the certificate with one stored on mobile device 100 and designated as trusted, or otherwise determined to be chained to a trusted certificate.

[0063] Mobile device 100 may also be adapted to store the private key of the public key/private key pair associated with the user, so that the user of mobile device 100 can sign outgoing messages composed on mobile device 100, and decrypt messages sent to the user encrypted with the user's public key. The private key may be downloaded to mobile device 100 from the user's computer 262a through cradle 264, for example. The private key is preferably exchanged between the computer 262a and mobile device 100 so that the user may share one identity and one method for accessing messages.

[0064] User computers 262a, 262b can obtain certificates from a number of sources, for storage on computers 262a, 262b and/or mobile devices (e.g. mobile device 100). These certificate sources may be private (e.g. dedicated for use within an organization) or public, may reside locally or remotely, and may be accessible from within an organization's private network or through the Internet, for example. In the example shown in FIG. 4, multiple PKI servers 280 associated with the organization reside on LAN 250. PKI servers 280 include a CA server 282 for issuing certificates, an LDAP server 284 used to search for and download certificates (e.g. for individuals within the organization), and an OCSP server 286 used to verify the revocation status of certificates.

[0065] Certificates may be retrieved from LDAP server 284 by a user computer 262a, for example, to be downloaded to mobile device 100 via cradle 264. However, in a variant implementation, LDAP server 284 may be accessed directly (i.e. "over the air" in this context) by mobile device 100, and mobile device 100 may search for and retrieve individual certificates through a mobile data server 288. Similarly, mobile data server 288 may be adapted to allow mobile device 100 to directly query OCSP server 286 to verify the revocation status of certificates.

[0066] It will be understood by persons skilled in the art that mobile data server 288 need not physically reside on a separate computing device from the other components

of LAN 250, and that mobile data server 288 may be provided on the same computing device as another component of LAN 250 in variant implementations. Furthermore, the functions of mobile data server 288 may be integrated with the functions of another component in LAN 250 (e.g. message management server 272) in variant implementations.

[0067] In variant implementations, only selected PKI servers 280 may be made accessible to mobile devices (e.g. allowing certificates to be downloaded only from a user's computer 262a, 262b, while allowing the revocation status of certificates to be checked from mobile device 100).

[0068] In variant implementations, certain PKI servers 280 may be made accessible only to mobile devices registered to particular users, as specified by an IT administrator, possibly in accordance with an IT policy, for example.

[0069] Other sources of certificates [not shown] may include a Windows certificate store, another secure certificate store on or outside LAN 250, and smart cards, for example.

[0070] Referring now to FIG. 6, a block diagram illustrating components of an example of an encoded message, as may be received by a message server (e.g. message server 268 of FIG. 4), is shown generally as 350. Encoded message 350 typically includes one or more of the following: a header portion 352, an encoded body portion 354, optionally one or more encoded attachments 356, one or more encrypted session keys 358, and signature and signature-related information 360. For example, header portion 352 typically includes addressing information such as "To", "From", and "CC" addresses, and may also include message length indicators, and sender encryption and signature scheme identifiers, for example. Actual message content normally includes a message body or data portion 354 and possibly one or more attachments 356, which may be encrypted by the sender using a session key. If a session key was used, it is typically encrypted for each intended recipient using the respective public key for each recipient, and included in the message at 358. If the message was signed, a signature and signature-related information 360 are also included. This may include the sender's certificate, for example.

[0071] The format for an encoded message as shown in FIG. 6 is provided by way of example only, and persons skilled in the art will understand that embodiments of the

invention will be applicable to encoded messages of other formats. Depending on the specific messaging scheme used, components of an encoded message may appear in a different order than shown in FIG. 6, and an encoded message may include fewer, additional, or different components, which may depend on whether the encoded message is encrypted, signed or both.

[0072] Embodiments of the invention are generally directed to a system and method for more efficiently searching certificates on a device and retrieving certificates for storage on the device. In one embodiment, the device is a mobile device (e.g. mobile device 100 of FIG. 4), and a certificate search application residing and executing on the mobile device is programmed to initiate searches of certificates on one or more certificate servers (e.g. LDAP server 284 of FIG. 4). In this embodiment, the mobile device searches for and retrieves individual certificates from a certificate server through an intermediate computing device (e.g. mobile data server 288 of FIG. 4).

[0073] With reference to FIG. 4, consider an example implementation where a certificate search application on mobile device 100 searches for and retrieves individual certificates from LDAP server 284 through mobile data server 288. A search request is received by the certificate search application, typically from a user who provides a first name, last name, and e-mail address of an individual for whom the user wishes to locate a certificate. Certain search requests may also be made broader, by constructing search queries where inputting only a few letters of a name will return all certificates issued with a name containing those letters as a prefix, or by otherwise using wild cards or blank entries in input fields to expand a search, for example. The search request is then communicated from mobile device 100 to mobile data server 288, which then queries LDAP server 284 for the requested certificate(s). In this example implementation, located certificates are retrieved by the mobile data server 288, and specific search result data relating to each retrieved certificate, such as the common name and e-mail address of the individual (or entity) to which the respective certificate is issued, is communicated to the mobile device 100 so that a list can be generated from the search result data for display to the user. The user can then select specific certificates for downloading to and storage on mobile device 100 from the list. The selections are then communicated to mobile data server 288, from which the selected certificates are downloaded to mobile device 100.

[0074] By communicating only specific search result data used to generate a list of located certificates instead of entire certificates to mobile device 100 in the first instance, and by only downloading specific certificates selected by the user, the search and retrieval of certificates can be performed more efficiently (e.g. in terms of time and bandwidth). However, prior art systems may not be adapted to determine or provide an indication to users as to which certificates on the list are already stored in a certificate store on mobile device 100, without downloading the certificates to mobile device 100 to facilitate a determination. In such systems, selected certificates may need to be downloaded in order to confirm that they are not already stored in the certificate store. This consumes time and bandwidth, and is potentially unnecessary.

[0075] Accordingly, embodiments of the invention relate generally to methods that can facilitate a determination of whether certificates may already be stored on a device (such as mobile device 100, for example), without requiring the certificates to be downloaded in their entirety to the device.

[0076] Referring to FIG. 7A, a flowchart illustrating steps in a method of searching and retrieving certificates in an embodiment of the invention is shown generally as 400.

[0077] At step 410, a first computing device receives a request from a second device to search at least one certificate server for certificates. In one example implementation, the first computing device acts as an intermediary between the second device and the at least one certificate server, such as a mobile data server (e.g. mobile data server 288 of FIG. 4) in the case where the second device is a mobile device (e.g. mobile device 100 of FIG. 4). In one example implementation, a certificate server to be searched may be an LDAP server (e.g. LDAP server 284 of FIG. 4).

[0078] The request may comprise data provided by a certificate search application executing and residing on the second device. The data may originate from user input to the certificate search application (e.g. when the search is initiated by a user), or from data generated by an application that initiates the search in variant implementations. The data will typically include at least one name and/or an e-mail address, although it will be understood by persons skilled in the art that a variety of search queries may be constructed without departing from the scope of the invention.

[0079] For convenience, further steps of method 400 will now be described with reference to an example implementation where the first computing device is a mobile data server and the second device is a mobile device. However, embodiments of the invention described with reference to method 400 or method 400b of FIG. 7B may be applied to implementations where the first computing device is not a mobile data server but is some other computing device and/or the second device is not a mobile device but is some other computing device. For example, a system architecture comprising first and second devices and at least one certificate server, where data transmissions between the first and second devices is generally more costly (e.g. in terms of time and/or bandwidth) than data transmissions between the first device and the at least one certificate server, may benefit from an application of an embodiment of the invention.

[0080] At step 420, the mobile data server queries the at least one certificate server for certificates based on the search request received at step 410 from the certificate search application of the mobile device. Certificates located in the search are retrieved from the at least one certificate server by the mobile data server.

[0081] At step 430, the mobile data server returns search result data relating to each located certificate to the certificate search application of the mobile device. The search result data returned typically includes the common name and e-mail address of the individual (or entity) to which the respective certificate is issued. However, in accordance with this embodiment of the invention, the mobile data server further processes each retrieved certificate, by parsing each retrieved certificate to identify the serial number and issuer of the respective certificate, which is returned as a part of the search result data.

[0082] In some implementations, the certificates retrieved at step 420 are only stored temporarily until search result data is returned to the mobile device at step 430, at which time the retrieved certificates are deleted. In other implementations, the certificates retrieved at step 420 may be cached or otherwise more permanently stored (e.g. until a response to the returned search result data is received from the mobile device, or for some pre-determined duration).

[0083] At step 440, the certificate search application compares the serial number and issuer data associated with each located certificate to the serial number and issuer data associated with certificates stored on the mobile device in one or more designated

certificate stores thereof, to determine whether the respective certificate is already stored on the mobile device.

[0084] At step 450, a list of located certificates is generated and displayed to a user of the mobile device. The list is generated from at least a subset of the search result data returned to the mobile device at step 430. For instance, the list may identify each located certificate by the common name and/or e-mail address of the individual (or entity) to which the certificate is issued. In one embodiment of the invention, an indicator accompanying the respective entry on the list for each located certificate may also be provided, where the indicator indicates whether the respective certificate is already stored on the mobile device, based on determinations made at step 440. Accordingly, the user need not select certificates for download that are already stored on the mobile device, so that duplicate certificates need not be necessarily downloaded to the mobile device. The indicator may comprise a checked or unchecked box, for example. As a further example, each entry on the list may be highlighted or non-highlighted depending on the status of the indicator.

[0085] At step 460, certificates are selected for download, by the user of the mobile device, for example.

[0086] At step 470, data identifying the selections made at step 460 are received by the mobile data server from the mobile device, and the selected certificates are subsequently returned to the mobile device, typically for storage on the mobile device. In some implementations, it may be necessary for the mobile data server to query a certificate server again for a selected certificate before the certificate is returned to the mobile device [step not shown], in the event that the certificate was not retained by the mobile data server since the previous download.

[0087] Referring now to FIG. 7B, a flowchart illustrating steps in a method of searching and retrieving certificates in another embodiment of the invention is shown generally as 400b. Method 400b is similar to method 400, except that the search result data associated with certificates returned by the first computing device to the second device comprises a hash of at least a part of each located certificate.

[0088] Specifically, at step 430b, the mobile data server returns search result data relating to each located certificate to the certificate search application of the mobile

device. The search result data returned typically includes the common name and e-mail address of the individual (or entity) to which the respective certificate is issued. In accordance with this embodiment of the invention, the mobile data server further processes each retrieved certificate by applying a hash algorithm, to hash at least a part of each retrieved certificate. The hash is then returned as a part of the search result data. In one implementation, the entire certificate is hashed to generate the returned hash. However in variant implementations, one or more specific parts or fields of a certificate may be hashed to generate the returned hash, although the likelihood that the hash will uniquely and correctly identify an identical certificate may be decreased depending on the parts or fields hashed.

[0089] At step 440, the certificate search application generates a hash for each certificate stored on the mobile device in one or more designated certificate stores thereof, and compares each generated hash to the hash associated with each located certificate, in order to determine whether the respective certificate is already stored on the mobile device. The same hash algorithm employed at step 430b is applied at this step (to the same parts or fields of the stored certificates if the entire certificates are not hashed) in generating the hashes of the stored certificates. Accordingly, if a generated hash of a given certificate matches a hash received from the mobile data server at step 430b, a match is deemed to have been determined.

[0090] Details pertaining to the remaining steps of method 400b are provided with reference to FIG. 7A.

[0091] In variant embodiments of the invention, other data that can be used to uniquely identify a certificate and that may be communicated more efficiently (e.g. in terms of time and/or bandwidth) than communicating the entire certificate may be returned to the second device as part of the search result data, and used to determine if the certificate is already stored on the second device.

[0092] Embodiments of the invention described above generally allow users to quickly determine which certificate(s) need to be downloaded to their computing device without making costly requests. In variant embodiments of the invention, the certificate search request may not be initiated by a user but may instead be initiated by an application executing on the second device (possibly by the certificate search application or some other application). In those embodiments, a list may not be generated for display to a user

(e.g. at step 450 of FIGS. 7A and 7B), and certificates may be automatically designated for download without user intervention after identifying which certificates may have already been stored on the second device (e.g. at step 440 of FIGS. 7A and 7B).

[0093] In variant embodiments, the invention may also be applied to other applications not involving certificates. For example, some of the foregoing techniques may be used to determine if certain contact data records or electronic documents are already stored on a computing device, for example.

[0094] The steps of a method of searching and retrieving certificates in embodiments of the invention may be provided as executable software instructions stored on computer-readable media, which may include transmission-type media.

[0095] The invention has been described with regard to a number of embodiments. However, it will be understood by persons skilled in the art that other variants and modifications may be made without departing from the scope of the invention as defined in the claims appended hereto.

Claims:

1. A method of searching and retrieving certificates comprising the steps of:
    a. receiving a certificate search request;
    b. performing a search on one or more certificate servers, wherein at least one query is submitted to the one or more certificate servers to request retrieval of certificates satisfying the certificate search request;
    c. retrieving at least one certificate from the one or more certificate servers;
    d. processing each certificate retrieved at step c) at a first computing device to determine search result data, wherein the search result data comprises data that uniquely identifies the respective certificate; and
    e. for each certificate retrieved at step c), communicating search result data associated therewith to a second device from the first computing device, for use in determining whether the respective certificate is stored on the second device.

2. The method of claim 1, wherein the data that uniquely identifies each respective certificate comprises serial number and issuer data for the respective certificate, and wherein the processing step comprises parsing each certificate retrieved at step c) to obtain the respective serial number and issuer data.

3. The method of claim 1, wherein the data that uniquely identifies each respective certificate comprises a hash of at least a part of the respective certificate, and wherein the processing step comprises applying a hash algorithm to each certificate retrieved at step c) to obtain the respective hash.

4. The method of claim 3, wherein the hash obtained for each certificate is a hash of the entire respective certificate.

5. The method of any of claims 1 to 4, wherein the second device is a mobile device.

6. The method of claim 5, wherein the first computing device is a mobile data server.

7. The method of any of claims 1 to 6, further comprising displaying a list of certificates retrieved at step c) to a user on the second device, wherein the list indicates which certificates are stored on the second device as determined at step e).

8. The method of claim 7, further comprising receiving one or more selections of certificates from the list, and communicating selected certificates to the second device.

9. A software application for execution on a computing device, the application comprising a plurality of instructions stored on a computer-readable medium, the instructions for implementing the method of any of claims 1 to 8.

10. A system for searching and retrieving certificates comprising at least a first computing device and a second device, wherein the first computing device is adapted to:

a) receive a certificate search request;

b) perform a search on one or more certificate servers by submitting at least one query thereto to request retrieval of certificates satisfying the certificate search request;

c) retrieve at least one certificate from the one or more certificate servers;

d) process each certificate retrieved to determine search result data, wherein the search result data comprises data that uniquely identifies the respective certificate; and

e) for each certificate retrieved, communicate search result data associated therewith to a second device for use in determining whether the respective certificate is stored on the second device.

11. The system of claim 10, wherein the data that uniquely identifies each respective certificate comprises serial number and issuer data for the respective certificate, and wherein in processing each certificate, the first computing device is adapted to parse each certificate retrieved to obtain the respective serial number and issuer data.

12. The system of claim 10, wherein the data that uniquely identifies each respective certificate comprises a hash of at least a part of the respective certificate, and wherein in processing each certificate, the first computing device is adapted to apply a hash algorithm to each certificate retrieved to obtain the respective hash.