Deployment of SIP in MANET: challenges and circumvention

Salma R. Abdelhamid1,2, Hossam-Eldeen M. Shamardan2, and Atef Z. Ghalwash2

1Faculty of Computers and Information Technology, Future University in Egypt, Cairo, Egypt
2Faculty of Computers and Information, Helwan University, Cairo, Egypt

Abstract. The immense growth in wireless network applications has encouraged the researchers to enhance and propose new approaches that facilitate the deployment of the widely used services, protocols, and applications of the wired networking area in the wireless networks. Among which, the Session Initiation Protocol (SIP) is a signaling protocol that allows the establishment of multimedia sessions and calls between different parties. SIP functionality totally depends on a centralized infrastructure, and complexity arises when deploying such a protocol in a special type of wireless networks, namely Mobile Ad-hoc Networks (MANET), due to the absence of a centralized infrastructure, and the dynamic topology of the latter. This paper highlights the challenges arising when deploying SIP in MANETs and presents an overview of different proposed approaches to overcome these challenges.

1 Introduction

The field of wireless networks has experienced an unprecedented growth during the past decades. The internet and Internet Technology (IT) devices such as laptops, Personal Digital Assistant (PDA), Internet Routers, and Smart Mobile phones play a vital role nowadays for a huge portion of persons. The presence of such a fast rate development of the used wireless networking technologies has persuaded the researchers to focus on an important type of networks, which is the MANET. The MANET is widely used over the globe due to the low cost of deployment and the flexibility of the network [1]. A MANET consists of a group of wireless mobile nodes (MNs) such as laptops, mobile devices, and PDAs connected together through wireless links. Unlike the traditional infra-structured wireless networks which mainly depend on centralized routers and fixed access points, MANETs are formed spontaneously and are mainly characterized with the absence of a centralized infrastructure [2]. In addition to that, the topology of the network faces rapid and frequent changes due to the mobility of the nodes, or sometimes their failures. These MNs of the MANET are self-organized and self-configured devices, and are able to communicate together despite this absence of an underlying structure or a centralized administrative support [3,4].

This de-centralized infrastructure-less attribute, directed the researchers towards proposing new approaches and modifications that facilitate the use of the world wide wired networks applications and protocols, one of which is the Voice over Internet Protocol (VoIP). The VoIP can be used to establish video or voice calls over the internet, taking over the role of the usual Public Switched Telephone Network (PSTN) [5]. When it comes to telephone communications, VoIP has proven to revolutionize this field, as one of its major advantages is saving end users from highly charged international calls. [6,7]. VoIP also provides additional features such as free long-distanced calls, determining caller ID, blocking certain callers or video-conferencing. However, the use of VoIP applications requires the establishment of a session between end users and for such establishment; the Session Initiation Protocol (SIP) is considered as a key element.

The Session Initiation Protocol (SIP) is a signaling protocol that is used to control multimedia communication sessions such as initiating, modifying, or terminating and interactive session. In other words, it is a signaling protocol that enables one party to call another party and to negotiate the parameters of a multimedia session. SIP is totally built upon a centralized infrastructure including
different entities such as Proxies and Registrars, typically owned by the network operator. Due to this major contradiction between the decentralized architecture of ad-hoc networks and the SIP, the later cannot be directly deployed into MANETs. And it is an undeniable fact that SIP multimedia services over MANETs can be indisputably exploited in different business areas, the thing that highlighted the exploration of its deployment in MANETs and different approaches were proposed [8]. This paper presents an overview of different approaches used for signaling SIP over MANET.

The rest of the paper is organized as follows. Section 2 presents an overview of the SIP, along with its entities and functionalities. MANET is briefly discussed in section 3, highlighting some of the routing protocols concepts. Section 4 wraps up the problem of deploying SIP in MANET. A literature review is presented in Section 5. Finally, the overall conclusion is summarized in Section 6.

2 SIP overview

Standardized by the Internet Engineering Task Force (IETF), SIP is a text-based and HTTP-like application-layer signaling protocol that is used for controlling multimedia communication sessions between end users [9]. It is independent on the underlying transport layer and is the most widely used protocol in wired networks IP-telephony. SIP is responsible of determining user location, confirming or denying user availability, and establishing and managing the sessions between users. These functions are totally dependent on a centralized architecture including Proxies and Registrars, typically owned by the network operator.

2.1 SIP architecture

SIP relies in its operation on a centralized architecture. Each entity is in charge of a certain function and means of communications between the entities are provided. The entities of SIP are divided into two main components, user agents and SIP servers [10].

2.1.1 SIP user agents

The SIP user agents are the end points in the network. They could be phones, gateways, or various types of computers. These end points exchange SIP messages and thereby manage a SIP session. The user agent can be a client, and thereby called User Agent Client (UAC) or can be a User Agent Server (UAS). The UAC initiates the session establishment requests and generates requests, such as REGISTER, INVITE, CANCEL, ACK, or BYE, whilst the UAS receives SIP requests and send SIP responses. [10].

2.1.2 SIP servers

SIP has four main servers, Proxy Servers, Registrars, Location Servers and Redirect servers [11]. Proxy servers are SIP intermediate entities. Depending on the dataflow; it could act both as a server and a client. Upon receiving the SIP request, the proxy server is responsible of selecting the best possible route to the desired UAS and forwarding the request to it. If the proxy server did not find the location of the targeted destination in its domain, it shall forward it to the responsible proxy server. It is also responsible for assuring the authentication of the end user and whether it is allowed to make a call or not. SIP Registrars are the servers at which SIP users should first register themselves. The users should provide their addresses, transport parameters, and their capabilities. Location Servers are the servers used by the registrars to store the locations of the registered users. And as obvious from their names, Redirect Servers are responsible of redirecting SIP users to another entity allowing them to connect to different set of addresses.

2.2 SIP functionality

The main role of SIP almost ends with the session establishment between the end users, and the actual data flow is directly carried out between the end users using the underlying transport layer, whether Transmission Control Protocol (TCP), or User Datagram Protocol (UDP). To identify SIP users, each of them is given a SIP address-of-record (AOR) Uniform Resource Identifier (URI). This address can be resolved at SIP proxy that is responsible for the users’ domain. Each user is supposed to register its address, using the SIP method REGISTER, at the SIP registrar in his domain, and thus allowing the identification of this user’s actual location in terms of IP addresses [12].

The nodes of the networks exchange different SIP messages, either requests or responses. When a node wants to initiate a connection, it sends an initiation request INVITE to its local proxy server without knowing the exact location of the called party. The proxy will determine the route to be taken to the callee after consulting the registrar and location server. The INVITE request is then forwarded by the proxy, and possibly through other number of proxies, until it gets to its destination. The callee could either accept or reject the incoming call. If accepted, the session initiation is then finalized by having the caller acknowledging the reception of the callees answer [10, 12]. After finalizing the session establishment, and as mentioned earlier, the data exchange between the two parties takes place without the involving of SIP proxy, and using the agreed upon transport protocol. Figure (1) illustrates a simple voice connection using SIP. The two involved end users are in different domains and have different proxies.
Fig. 1, Setting up a voice call using SIP [13]

The numbered steps are explained as follows [13]:
(1) User Agent “B” registers itself at its local SIP registrar.
(2) The Registrar stores the location at the location server.
(3) User Agent “A” sends an INVITE request to its local Proxy.
(4) The Proxy of user A translates the destination address domain and forwards the request to the proxy of user B domain.
(5) Proxy in Domain B looks up the address of user B in the location server.
(6) The Location server replies with the required IP-Address.
(7) Proxy in Domain B sends the request to the desired user agent.
(8-10) The response message from A to B takes the reverse route back and then after finalizing the session establishment the media content is to be exchanged directly between the end points.

3 MANET overview

The MANET is a collection of two or more devices or nodes or terminals with wireless communications and networking capability that communicate with each other without the aid of any centralized administrator [14].

3.1 MANET characteristics

Unlike traditional wireless networks, the nodes of a MANET do not have fixed locations, instead, they move from one position to another causing an unpredicted dynamic change in the topology of the network. Also besides the nodes mobility within the network, MANET nodes can anytime leave the network completely due to their movement or failure. In addition to that, one of the challenges that exist in MANET is the limitation of the available bandwidth. The MANET nodes do not have the privilege of transmitting within a large bandwidth as the devices in the wired networks. Moreover, the fading of the signal, noise effect and signal interference degrade the wireless link capacities [3, 14].

3.2 MANET routing protocols

As a result of the challenging MANET characteristics, the protocols designed for centralized wired or wireless networks cannot be directly applied to MANET. The end users of ad-hoc networks compensate the decentralized infrastructure-less architecture by relying on each other in routing. Each node contributes in routes discovery, and data sending from the source to the desired destination usually relies on multiple hops.

MANET routing protocols take into consideration the dynamic topology of the network, as well as the absence of a fixed network infrastructure [15-17]. And for the protocol to be effective, lots of operations and properties should be supported, amongst which are reactive and demand based operations, distributed operations, multiple routes providence, security measures and looping avoidance [16-18]. A lot of classifications exist for Ad-hoc networks Routing Protocols [18]. The most common classification is based upon how routing information is acquired and maintained by the nodes of the network. Accordingly, the Ad-hoc routing protocols can be divided into three types: Proactive, Reactive, and Hybrid routing protocols [4, 2, 15, 16, 19].

3.2.1 Proactive (Table-Driven) routing protocols

In this type of protocols, each node maintains the routes to all possible reachable destinations of the network in its routing table, and these tables are periodically updated by propagating update messages with the help of a sequence number created by the destination to disseminate fresh routes in case any change in the network topology occurred. The node should keep up-to-date routing information so that the packet is directly forwarded when required to the destination. The routing protocols based on this methodology differ in the way nodes update the topology change of the network and the routing information maintained in the routing tables.

However, this type of protocols does not perform effectively in highly dynamic networks as for each topology change; the MNs are obliged to update their routing tables causing an increase in the control message overheads thus leading to an overall network performance degradation [4]. Examples of this protocol are Destination sequence Distance Vector (DSDV) [14, 20, 21], Wireless Routing Protocol (WRP) [22, 14] and Clusterhead Gateway Switch Routing [14, 18, 23].

3.2.2 Reactive (On-Demand) routing protocols

In this type of protocols, each node maintains the routes to all possible reachable destinations of the network in its routing table, and these tables are periodically updated by propagating update messages with the help of a sequence number created by the destination to disseminate fresh routes in case any change in the network topology occurred. The node should keep up-to-date routing information so that the packet is directly forwarded when required to the destination. The routing protocols based on this methodology differ in the way nodes update the topology change of the network and the routing information maintained in the routing tables.

However, this type of protocols does not perform effectively in highly dynamic networks as for each topology change; the MNs are obliged to update their routing tables causing an increase in the control message overheads thus leading to an overall network performance degradation [4]. Examples of this protocol are Destination sequence Distance Vector (DSDV) [14, 20, 21], Wireless Routing Protocol (WRP) [22, 14] and Clusterhead Gateway Switch Routing [14, 18, 23].
to a certain destination. When a node wants to send a packet to a certain destination, it applies the route discovery mechanism which mainly depends on the idea of flooding the network with route query requests to this destination [4]. The determined route remains valid until the destination is reached or any link breakage occurs in it. The drawback of the reactive protocols is the delay caused during the route discovery process which makes it inconvenient for time sensitive applications. An example of this protocol is the Ad hoc On-demand Distance Vector Routing (AODV) [16, 24, 25].

3.2.3 Hybrid routing protocols

The hybrid protocol is a merge between both proactive and reactive protocols. The routing is initially established using the proactive approach, and then the route requests from additionally added nodes are provided using the reactive approach. This approach tends to reduce the overhead of the proactive protocols as well as the time delay of the reactive protocols [4]. An example to such protocol is the Zone Routing Protocol (ZRP) [15, 26].

4 Problem definition

Ad-hoc networks have a great advantage of designing flexible networks. However, as mentioned earlier, the mobility of the nodes and the lack of a centralized infrastructure remains a challenging constraint in some research areas. Due to such constraints, the nodes are not familiar with the topology of their networks. Instead, they have to enter a discovery phase first. When joining the network, each node should be able to announce itself and discover neighboring nodes to be able to send its data to the required destination. Afterwards, Routes should be discovered and transmission takes place using cooperative multi-hops between source and destination. Another key problem with ad-hoc networks is that their dynamic un-centralized architecture makes it difficult to implement any form of communication requiring sessions instead of the best-effort dissemination of independent packets from and to multiple nodes.

SIP on the controversy to Ad-hoc networks requires centralized entities, such as the Proxy server, the Registrar or the Location Server. Each one of SIP servers has its own specific role in setting up the connection and defining the routes between sources and destinations. This centralized SIP architecture is obviously not straightforwardly applicable to ad-hoc networks. As mentioned, ad-hoc networks are wireless dynamic networks formed by peer nodes while SIP proxies and registrars are static, fixed, and centralized entities. SIP users in MANETs cannot reach other parties, as they do not have support from proxy servers. Moreover, they cannot be reached by other nodes, as there are no SIP registrars where they can register their contact information. Nevertheless, the Quality of Service (QoS) metrics of VoIP such as bandwidth consumption, delays, jitters, packets loss, and signaling are severely degraded as the traffic and number of multiple hops to the gateway increase. Consequently, several problems arise when directly deploying SIP services in ad-hoc networks, and altering the SIP main centralized architecture should first take place.

5 Deployment of SIP in MANET

This section presents an overview of some of the existing literature on researches addressing SIP signaling over MANET.

The main idea proposed by Leggio et al. in [27] is to decentralize SIP to be able to deploy it in MANETs. Decentralizing was fulfilled by embedding a limited set of SIP server functionalities into SIP user agents. Thus, the support originally provided by centralized SIP entities is merged in all end users in the ad-hoc networks, and each node carries out the required function in order to establish the session.

The authors present two methods to discover the users’ AOR and thus allow session establishment. Those two methods are Fully Distributed SIP (dSIP) and SIP with Service Location Framework (sSIP). dSIP is to dispense the need of centralized SIP entities, but rely on SIP methods to acquire user contact information. When a broadcasted REGISTER message is received, it is stored in the local cache of the node, and when an INVITE message is received, the local proxy modules plays the role of binding between the specified address and SIP AOR. sSIP on the other hand depends on Service Location Protocol (SLP) [28]. In such a case, each node receiving the broadcasted SLP request replies back with SLP reply including its binding. Scalability is not an advantage of this approach due to the message broadcasting which leads to bandwidth consumption and flooding in large MANETs.

Stuedi et al. [29] handled the setup and management of the sessions in MANETs using SIPHoc, a middleware infrastructure. SIPHoc is completely decentralized, which eliminated the need of centralized components, or assigning special functionalities to some elected nodes of the network. Independent on the underlying network topology, SIPHoc supports mobile and static ad hoc networks, and it does not require any modifications in the existing applications. Moreover, this approach is based upon MANET SLP for registration and look-up operations which works by piggybacking techniques that provides an efficient messaging system.

L. Chang et al. in [30] embedded a server in the transport layer of each MN called pseudo Session Initiation Protocol (p-SIP) server. This server is meant to provide VoIP services over ad-hoc networks. The p-SIP includes the user discovery module, the session management module, the user-list cache, the SIP presence module, and
Moreover, the problem of the nodes mobility is resolved by means of an additional introduced mobility management module that allows each node to maintain the intended SIP information. The ad-hoc discovery mechanisms are handled via standard SIP protocol. The implementation of p-SIP is done on real devices, and the performance measurements prove the feasibility to realize the VoIP services over ad-hoc using the designed p-SIP server.

H. Chu and W. Chen proposed in [31] an integrated middleware to play the role of SIP servers and takes over their responsibilities and functions. The client registration step is to take place at this middleware instead of SIP servers. The used underlying routing protocol is AODV, which delivers the exchanged SIP messages. The authors refined the registration process by forcing the middleware to act as a registrar server, thus locally binding the IP address with the AOR. The process of forwarding the SIP messages which is supposed to be performed by the SIP proxy server is also carried out by the proposed middleware. The goal of this approach was enabling SIP applications in MANETs in addition to minimizing the signaling overhead in the network.

Mourtaji et al. proposed in [32] building a Virtual Network for Session Initiation Protocol (VNSIP). The main idea was to self-organize the Ad-hoc Network using a virtual backbone. The virtual network was constructed using specific nodes within the ad-hoc network. The SIP servers’ functionalities are to be embedded into all nodes of the MANET, and according to a Virtual Network Algorithm (VNA); each node gets the function of a certain SIP server activated so that at the end, all SIP servers would be included in the network to accomplish the address distribution task. During the registration process and depending on whether the node is a member of the VN or not, the SIP are either locally sent to the node itself or broadcasted to all of the 1-hop neighboring nodes. This broadcasting is known as the Replication Mechanism. The replication mechanism allows searching in multiple SIP Proxies at the same time, and thus ensuring short response time. The simulations prove that the performance of VNSIP was remarkable at the session establishment time. The simulations prove that the performance of the previously mentioned VNSIP [32], Mourtaji et al. proposed an enhancement later in [33]. The problem mainly arises because it is not possible to simultaneously accept all communications due to MANET bandwidth limitations. The proposed algorithm, MANET Call Admission Control (MCAC), is used to permit the establishment of calls to a number that agrees with the available bandwidth of the MANET, otherwise it rejects the calls by a certain SIP error code resulting in a busy tone. The decision of establishing or denying communications is based upon a specific threshold which is related to the network gateways responsible of providing connectivity to the MANET. When this threshold is exceeded, the establishment of new sessions is refused, and thus the already established communications remain undisturbed.

Almobaideen et al. presented an application layer called; Fuzzy and Cluster based SIP Protocol (FCSIP) [34]. FCSIP is independent on the underlying routing protocol and the MANET in which this protocol is applied is assumed to be clustered. The main idea of clustering is to divide the nodes of the network into virtual zones called cluster. Each cluster comprises a group leader called Cluster Head. The cluster head is responsible for managing its cluster and coordinate with other clusters in the network. Other members of each cluster are either cluster gateways which have access to neighboring clusters and are used to forward information between them or cluster members which are neither CH nor gateways [18, 35].

Aside from several used algorithms for electing the CH [36], the authors define a couple of criteria in their fuzzy algorithm, one of which resembles the VoIP activeness of this elected node. The main advantage of using an underlying cluster-based routing protocol is the significant reduction of transmitted SIP control messages, and this comes as a result of the hierarchical level represented by the CH and the members falling beneath it. The enhancement is proven by the simulation that compared FCSIP to a fully distributed version of SIP which they refer to as FD SIP.

6 Conclusion

Being a worldwide used protocol for supporting VoIP applications, the inherently centralized SIP encouraged the researchers to intentionally expand in proposing new approaches to facilitate the deployment of SIP in wireless networks. This paper presented a survey on the implementation of SIP functionality over a simultaneously formed and a decentralized wireless network such as the MANET. These implementations require a lot of adaptations and amendments. Some researchers proposed fully decentralizing the SIP functionalities while others proposed a middleware infrastructure to play the role of SIP servers. And incontestably, the future work in this field of interest is highly promising.

References