ABSTRACT

Ad hoc networks have lots of applications; however, a vital problem concerning their security aspects must be solved in order to realize these applications. The dynamic and cooperative nature of ad hoc networks present challenges in securing these networks. There are recent research efforts in securing ad hoc networks. Amongst security approaches, there are threshold cryptography, certification authority, reputation and authentication. In this paper we introduce and survey these approaches. We conclude this paper and identify the challenges and open research areas associated with each of these approaches.

Keywords: Ad hoc networks, authentication, certification, reputation, threshold cryptography, security.

1 INTRODUCTION

Mobile ad hoc networks are generally characterized by the lack of infrastructure, dynamic network topology, distributed operation, bandwidth constraints, variable capacity links, use of low power devices, limited CPU and memory, limited physical security, and complexity of design of network protocols. However, ad hoc wireless networks are highly appealing for many reasons. The set of applications for mobile ad hoc networks is diverse, ranging from small, static networks that are constrained by power sources, to large-scale, mobile, highly dynamic networks. The dynamic and cooperative nature of ad hoc networks present challenges in securing these networks. There are recent research efforts in securing ad hoc networks. Amongst security approaches, there are threshold cryptography, certification authority, reputation and authentication.

In this paper we survey those approaches and identify the challenges associated with each. The remainder of this paper is organized as follows. Section 2 is concerned with the threshold cryptography based schemes whereas section 3 focuses on the certification authority schemes. In section 4 trust and reputation based schemes are presented and in section 5 authentication schemes are surveyed. Finally, conclusions and future challenges are given in section 6.

2 THRESHOLD CRYPTOGRAPHY

In this section we survey different threshold cryptography schemes proposed for ad hoc networks and the solutions suggested in the literature for determining the optimum threshold level. This will be presented in sections 2.1 and 2.2 respectively.

2.1 Threshold Cryptography Schemes

Security schemes for ad hoc networks generally use public-private key mechanism. The overall system has a known public key and its private key is shared by between each server nodes in the system. Each server node stores the public key of other elements and sign request responses using the private key of the overall system. Requests may be update the node’s public key or query the public key of the node that is intended for private communication. New public key of the node can be broadcasted since combiner should use the private key of the server system to obtain it. System is secure because
adversary does not have enough computational power to break these cryptographic schemes; it is also robust that servers are always able to process update and query requests. Threshold cryptography is the base stone for distribution of trust protocols. The idea of \((k, n)\) threshold scheme was introduced by Shamir in [1]. A \((k, n)\) scheme allows a secret, to be split into shares, such that for a certain threshold \(k=n\), any \(k\) components could combine and generate a valid signature; whereas, \(k-1\) or fewer shares are unable to do so. Zhou and Haas in [2], proposed the idea of utilizing threshold cryptography to distribute trust in ad hoc networks. According to [2], the challenges associated with key management services such as issuing, revoking and storing of certificates in ad hoc networks can be resolved by distributing Certification Authority (CA) duties amongst the network nodes.

In [3] a hierarchical Public Key Infrastructure (PKI) was suggested for ad hoc networks. In this scheme, distribution of trust is achieved using threshold cryptography. Some threshold schemes exploit redundancies in the partial signatures and use error correcting codes to mask incorrect partial signatures [4]. With these schemes a correct signature is obtained despite a small number of partial signatures being incorrect, this means that the scheme will recover from corrupted nodes that return incorrect partial signatures. Link keys are established using flooding. All connected nodes broadcast their own signed wake up call to neighbors and so on. Secure authentication of wake up calls is used to avoid replay and battery drain attacks, and dynamic behavior to reestablish broken chains. Figure 1 depicts the suggested PKI hierarchy.

<table>
<thead>
<tr>
<th>layer</th>
<th>certification keys</th>
<th>certificates</th>
</tr>
</thead>
<tbody>
<tr>
<td>level 0</td>
<td>SK_0 = (s_1, ..., s_n)</td>
<td>Cert_{0}(PK_{1,1}) ... Cert_{0}(PK_{l,m})</td>
</tr>
<tr>
<td>level 1</td>
<td>SK_1 = (s_1, ..., s_k)</td>
<td>Cert_{1}(PK_{2,1}) ... Cert_{1}(PK_{2,n})</td>
</tr>
<tr>
<td>level 2</td>
<td>SK_2 = (s_1, ..., s_k)</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>level m</td>
<td>...</td>
<td>Cert_{m-1}(PK_{m,1}) ... Cert_{m-1}(PK_{m,k})</td>
</tr>
</tbody>
</table>

Figure 1: Distributed and hierarchical PKI for ad hoc networks. Shares of the private key SKI at layer 1 are signed by the private key SKI−1 of layer 1 − 1 [3]

As it is shown in Figure 1, on the top layer of the hierarchy a master certification SK is used to issue certificates for the public keys of the nodes on level 1. Next to this, all nodes on level 1 get a share of the layer 1 certification key SK1. Similarly, level 2 nodes receive a certificate signed by SK1 and a share of the layer 2 certification key SK2. This process is continued until the desired number of levels in the hierarchy is reached.

Distribution of trust is achieved using threshold cryptography [5], [6]. An \((n, k + 1)\) threshold scheme that allows \(n\) parties to share the ability to create a digital signature is used, so that any \(k + 1\) parties can perform this operation jointly, whereas it is infeasible for at most \(k\) parties to do so. The certification keys SKi are divided into \(n\) shares \((s_{i1}, s_{i2}, ..., s_{in})\). If a node at some level requires a certificate, it will contact \(k + 1\) nodes of the previous level (up) to gather \(k + 1\) partial signatures and combine them to compute the signature for the certificate.

An Anonymous and Certificateless Public-Key Infrastructure (AC-PKI) to efficiently and securely provide public-key services without using public-key certificates was proposed in [7]. To satisfy the demand of private keys during network operation, a distributed private key-generation scheme was designed by utilizing Shamir’s \((k, n)\) secret sharing technique to distribute a system master-key among a set of pre-selected nodes, called Distributed Public Key Generators (D-PKGs). In addition, D-PKGs were offered anonymity protection to defend against pinpoint attacks, which makes AC-PKI more secure than previous applications of the secret-sharing technique in mobile ad hoc networks [7].

In [8], a new scheme based on which the verifiability is achieved in a simple manner was presented. It controls the joining of a node in the network to give it a share to make it able to participate in accepting other nodes. To control admission to a secure group, the general membership model control is as follows:

- Setup: In the initial phase, each group member obtains his secret share and a group membership certificate (GMC) from an offline-centralized dealer or by collaborative computation among initial group members.

- GMC Issuance: A prospective member initiates the protocol by sending a join request message to the group. If \(k\) members or more approve of admission, they will cooperatively generate the GMC of the prospective member.

- Share Acquisition: If the new member becomes a legitimate member, he acquires his own share which enables him to participate in future admission protocols.

In [9], a secure and effective distributed certification service method was proposed using the secret sharing scheme and the threshold digital signature. In the proposed distributed certification service, certain nodes of relatively high safety among the mobile nodes were set as privileged nodes, from which the process of issuing a certification starts. The proposed scheme solved problems that would have damaged the whole network security by the intrusion of one node in the centralized architecture and the hierarchical architecture. Also, it decreases the risk of exposure of the private keys in the fully distributed architecture as the number of the nodes containing the partial confidential information of personal keys decreased.
A mechanism that allows creation of a keying service in the network was suggested in [10]. It is a novel combination of two cryptographic techniques: ID based and threshold cryptography. ID-based cryptography primarily provides efficiency gains, and threshold cryptography provides resilience and robustness. Particular schemes were identified as candidates for implementing this approach. However, the scheme is vulnerable to man in the middle attacks on joining members.

2.2 Optimum Threshold Level
If threshold cryptography is used, it is important to know the value of the threshold k. A very high threshold level ensures greater security, but the QoS requirement may not be satisfied. If the threshold level is lowered, it becomes easy for a node to construct its digital certificate within the QoS requirements or specified authentication delay time, but the security aspect is compromised. The threshold level selection process is influenced by various network dynamics such as network density, node speed, node transmission range, threshold requirements etc. In [11], the calculation of the threshold level was modeled as an optimization problem for a certain QoS requirement. However, this optimization problem cannot be solved with standard optimization techniques as the function is not known. Therefore, simulations were used to optimize the threshold level function to derive the optimum threshold level. Two ways were investigated to fix the threshold level. First method: Global Selection, where the threshold level is fixed, i.e., it is the same for all nodes at all times. Second method: Local Selection, where the threshold level is selected based on the local environment of a node at that moment. This method is more responsive to the dynamic nature of a mobile network. The results have shown that in global selection protocol, the biggest drawback is that the number of partial certificates required to construct the full certificate is fixed for all the nodes in the network. This results in failure to construct certificates as the QoS requirements cannot be met. According to [11], the network traffic increases steeply as a result of the higher number of certificate construction failures; this could result in network congestion. In local selection protocol, the required number of partial certificates is determined based on the locality of a node. Moreover, it is easier to select the critical threshold value for a given network. However, due to more number of steps involved in the protocol, performance of the protocol drops down for nodes that move at higher speeds. But this can be overcome by setting precedence level to certificate request packets. An intelligent approach to determine the optimum threshold level given a network configuration using neural networks was also proposed in [11]. A trained neural network can be embedded into each node, so that nodes can compute an optimum threshold level for different network conditions and use it in the authentication protocol.

In [7], the optimal secret-sharing parameters (k, n) were calculated to achieve the maximum security and a novel protocol was designed to dynamically adjust (k, n) to accommodate dynamic node join, leave. If $P_{comp}$ is the probability that adversaries happen to pick up and compromise k Distributed Private Key Generators D-PKGs in one time period so as to reconstruct the system master-key, and $P_{para}$ the probability that adversaries happen to pick up (n-k+1) D-PKGs and corrupt them in one time period so that there are no enough k D-PKGs to collaboratively provide the requisite private key generation PKG service, the following equations were obtained in [7]:

$$\Pr_{comp} = \frac{n}{N} \prod_{i=k}^{N-i} \frac{k}{n}$$  \hspace{1cm} (1)

$$\Pr_{para} = \frac{n-k+1}{N} \prod_{j=0}^{n-j} \frac{n-j}{N-j}$$  \hspace{1cm} (2)

Where N and n denote the numbers of nodes and D-PKGS in the network, respectively. In practice, both metrics are equally important and expected to be as low as possible. To reflect this fact, a new metric Security Level (SL) was calculated.

$$SL_n(k) = 1 - 0.5 \times \Pr_{comp} - 0.5 \times \Pr_{para}$$  \hspace{1cm} (3)

3 CERTIFICATION AUTHORITIES
In order to have threshold cryptography, certification authorities (CAs) are needed. This section focuses on CAs. The concept and tasks of the CAs is presented in section 3.1, and a comparison between the single and multiple CAs case is given in section 3.2. In section 3.3 the certification schemes in ad hoc networks are given, whereas in section 3.4 the certificate revocation schemes are presented.

3.1 Concept of Certification Authorities
In ad hoc networks, trust is managed locally at the individual nodes. A node is not trusted by a given node until it presents a certificate, and the node in question verifies that the certificate was issued by a trusted CA, and it has not expired nor been revoked. The CAs have the following trust management tasks [12]:

1) Issuing of certificates
2) Storage of certificates
3) Certificate validation
4) Revocation of certificates.

Beyond managing certificates, it is also the CA’s responsibility to disseminate the public keys of principals to inquiring clients. Every response from the CA is signed with the CA’s private key, and so can be validated with the CA’s public key. The success of this approach lies in maintaining the secrecy of the private key of the CA. It is also necessary for the CA to remain on-line (i.e. available) to provide these services.

There are three major parameters to a distributed key management framework: fault tolerance, vulnerability and availability. The first parameter is associated with the number of node failures the system can handle, the second is associated with the number of compromised nodes the system can withstand, whereas the third is associated with the ability of the client to contact the required number of CAs. The optimization of any one of these parameters may adversely affect other parameters and so adversely affect the success of the system. In addition, mobile networks present hostile environments where nodes may easily die or be compromised and no guarantees can be made about the ability to access the necessary nodes for authentication. An ideal key management service for ad hoc networks should provide the best of both worlds: it must be light-weight and simple to mobile nodes, and it must be available in highly dynamic networks.

3.2 Certification Authorities Selection

A single centralized authentication server is unsuitable for ad hoc networks, from the security point of view, as it may be subject to a single point attack. To provide better fault tolerance, it is possible to deploy many copies of the CA in the network.

Different certification schemes have been presented in the literature. We classify these schemes into cluster-based schemes and non-cluster-based schemes and present them in subsections 3.3.1 and 3.3.2 respectively.

3.3 Certification Schemes in Ad Hoc Networks

In a cluster-based architecture for a distributed public key infrastructure that is highly adapted to the characteristics of ad hoc networks was introduced in [16]. In order to adapt to the highly dynamic topology and varying link qualities in ad hoc networks, central instances that would form single points of attack and failure were avoided. Instead, the ad hoc network was divided into clusters, and the cluster heads jointly perform the tasks of a certification authority. A proactive secret sharing scheme distributes the private network key to the cluster heads in the ad hoc network. Instead of a registration authority, arbitrary nodes with respective warranty certificates may warrant for a new node’s identity. Based upon this authentication infrastructure, a multi level security model ensuring authentication, integrity, and confidentiality is provided. Authentication itself is realized in two stages. First, a node gets the status of a guest node.
After sufficient authentication, the node will become a full member. An additional important feature is the possibility to delegate the cluster head functionality to another node. [16]

Another approach based on trust model and clustering algorithm was proposed in [17] in order to distribute a CA. The clustering algorithm is based on two parameters, security and stability. The security factor is related to the trust model; only confident nodes can become cluster-head and assure CA role. In each cluster, there are five roles of nodes: The CA Certification Authority of cluster k which certifies public key of nodes belonging to the same cluster, the RA Registration Authority which protects CA against attackers. The GW is a gateway node ensuring a connection between two different clusters i and j, these nodes must be certified by two different CAs. The MN represents a member node i which belongs to the cluster k. Finally, the VN is a visitor node i that belongs to cluster k, it has low trust certificate. In the clustering algorithm, the stability factor is presented by mobility metric in order to give more stable clusters. The trust model is evolved by monitoring process which allows any node with high trust metric to monitor and evaluate other nodes with low trust metric. To protect CA nodes, a Dynamic Demilitarized Zone (DDMZ) permits to increase security robustness of cluster and endure malicious nodes that try to attack CA or issue false certificates. This approach ensures the security and availability of public key authentication in each cluster and this architecture is adapted to any topology changes.

An Authentication Service Based on Trust and Clustering in Wireless Ad Hoc Networks was described and evaluated in [18]. It is a combined reputation and authentication scheme in which there are two types of trust: direct within same cluster and recommended between different clusters. For certification within the same cluster, there is no problem as nodes know each other. For certification within different groups, the node selects n nodes (called introducers) with the highest trust values and sends them request messages. Before sending out the request message, node vi first checks whether it is in the same cluster as vj. If it is, it sends the request message to its neighboring nodes, assuring that some of its neighboring nodes have built up a direct trust relationship with vj. On the other hand, if vi and vj are in different clusters, then the problem becomes more complicated. Node vi has to select some trustworthy nodes in the target cluster to be the introducing nodes, or so-called introducers, they are nodes in the same cluster as vj for which vi has high trust values. However, it is possible for the introducers to be malicious; therefore, a voting procedure is carried out to conclude the correct public key of the target node by majority vote. Identification and isolation of malicious nodes is done using three methods. First Method: Direct monitoring of individual nodes by listening to the traffic via wireless communications using a monitoring facility. Second: By identifying suspicious introducers who provide public key certificates different from the others. Third: If the trust values provided by the introducer indicate a node is malicious. To deal with colluding nodes a scheme is suggested. After filtering out suspicious introducers, the trust value of a target node t is obtained from the rest of introducers.

3.3.2 Non Cluster-Based Certification Schemes

In [15], a certification protocol called MP (MOCA Certification Protocol) was proposed. Given the threshold value, k, the total number of nodes, M, and the number of MOCA servers is one to (k or more) then back, which means that a client needs to contact at least k MOCA servers and receive replies from each of them. To provide an efficient way of achieving this goal, a certification protocol called MP (MOCA certification Protocol) was proposed in [15]. In MP, a client that requires certification services sends Certification Request (CREQ) packets. Any MOCA that receives a CREQ responds with a Certification Reply (CREP) packet containing its partial signature. The client waits a fixed period of time for k such CREPs. When the client collects k valid CREPs, the client can reconstruct the full signature and the certification request succeeds. If too few CREPs are received, the client’s CREQ timer expires and the certification request fails. The client is left with the option to initiate another round of certification requests.

As a CREQ packet passes through a node, a reverse path to the sender is established. These reverse paths are coupled with timers and maintained long enough for a returning CREP packet to be able to travel back to the sender. The management of routing information in the intermediate nodes and the use of reverse path forwarding of CREP packets are similar to on-demand ad hoc routing protocols like AODV [19] or DSR [20]. While the use of flooding approach to reach all MOCA is effective, it generates quite a large amount of overhead traffic. First, the traffic generated from CREQ flooding is large. Second, since a client has no way to limit the dissemination of a CREQ, all the MOCA that receive a copy of the CREQ respond with a CREP, making the client receive more than it actually needs to reconstruct the full signature. Note that a client only needs to collect k partial signatures to reconstruct the full signature. Any additional partial signatures are discarded and waste networking and processing recourses. To reduce the amount of overhead from the flooding while maintaining an acceptable level of service, another method called f-unicast was introduced. In f-unicast, if a client has sufficient routes to MOCA’s in its routing table, the client can use multiple unicast connections to replace
flooding. This scheme takes advantage of existing routes, as seen in the routing table. Blind use of unicast with insufficient cached routes can result in multiple instances of route discovery, which in turn causes multiple rounds of flooding. To prevent such a situation, the protocol only uses unicast when there is enough information cached in the routing table; otherwise it falls back to flooding.

\(a\) is a marginal safety value to increase the success ratio of unicast and it should be determined based on the node’s perception of the network status. But if there is more than the sufficient number of routes in the cache, the choice of which ones to use can affect performance. Three different schemes were defined:

- Random MOCAs: Random \(k + a\) MOCAs in the routing table are picked.
- Closest MOCAs: By using the hop count information stored in the routing table, \(k + a\) MOCAs with smallest hop counts are used in this scheme. Intuitively, this approach has the benefit of the shortest response time and the smallest packet overhead since the CREQ packets travel the least distance.
- Freshest MOCAs: Among the MOCAs in the routing table, the most recently added or updated \(k + a\) entries are used for \(\beta\)-unicast. This approach is least vulnerable to possible stale routes in the route cache, especially under high mobility.

Another framework for a distributed KMS that increased service availability for highly partitioned networks was proposed in [21]. The system integrated a number of components in a unique way to counteract the limitations of previous KMSs. As it is shown in Figure 2. The system utilized a modified hierarchical PKI model consisting of a control plane of Root Certification Authorities RCAs, Delegate Certification Authorities DCAs, and Temporary Certification Authorities TCAs. The RCAs authenticated new nodes and issued them RCA certificates. New nodes could use the RCA certificates to register in the network and serve as DCAs, minimizing pre-configuration. In addition, new nodes could establish temporary Security Associations SAs in the absence of DCAs, thus introducing more flexibility into the KMS. The DCAs issued, revoked, distributed and managed certificates based on the behavior grading of the nodes and the security policies at the network and node level. The TCAs aided new nodes to join the network by issuing temporary certificates whenever DCAs were unavailable. In addition, the Trusted Peers TPs of each node acted as repositories increasing the availability of certificates in a partitioned network. In addition to revocation [22], security in the KMS was provided via behavior grading and non repudiation.

In [23], Distributed Certification Authority with probabilistic freshness for Ad Hoc Networks (DICTATE) was proposed. It focused on the design of a certification authority in ad hoc networks. It consists of a joint authority approach that combines an offline identification authority and an online distributed revocation authority. Authority consists of mCA (mother CA) & distributed CA (dCA) node servers & the clients.

dCA has a public private key pair. Public key issued by mCA known to the whole network & the private key is shared among dCA servers by a robust threshold cryptosystem. Periodically, there is a check time at which the servers (physically) go back to the mCA for a purgation for mCA to detect compromised servers and substitute them. The key revocation of a server is done by using a public key & the combination of identity & the time stamp corresponding to a certain check time interval. A client can verify the validity of a message form any server using its ID and a local clock loosely synchronized with the check time. The Operating principle of the joint CA at the initialization phase and the check time as described in [23] are depicted in Figure 3 and Figure 4 respectively.
3.4 Certificate Revocation Schemes

In this section, we focus on the certificate revocation in ad hoc networks. The importance of the certificate issue is discussed, and the certificate revocation schemes that have been used for ad hoc networks are presented. Of all trust management tasks, certificate revocation poses the most challenges [12]. For various reasons, certificates will need to be revoked periodically; for example, if the private key associated with a certificate is compromised, the certificate will need to be revoked and information be made available to network peers in a timely manner [25].

Certificate revocation is an issue too important to be ignored; nonetheless, if adequate safeguards are not put into the process of determining when a certificate should be revoked, malicious nodes can wrongfully accuse other nodes of misbehavior and cause the certificates of good, uncompromised nodes to be revoked. Compromised or malicious nodes can in fact use this phenomenon (we called it malicious accusation) as an exploit for isolating and ultimately cutting off legitimate, well-behaving nodes from a network [25].

For traditional networks with online access to centralized repositories or CAs, revoked certificates are usually declared in certificate revocation lists (CRLs) [26], and the CRLs are either placed in easily accessible repositories, or broadcasted to the relevant nodes [12]. Alternatively, online certificate status protocol (OCSP) [27] can be used to ascertain information about the status of a certificate [25].

In [23] a revocation scheme was suggested together with the certification scheme. Periodically, there is a checktime, at which the distributed CA (dCA) servers (physically) go back to the mother CAs (mCA) for purgation (only distributed CA servers should go through this procedure; clients can still perform their remote operations). During the checktime, the mCA, through out-of-band mechanisms, detects compromised servers and has them reinitialized or substituted by new ones; it also refreshes the secret shared among the dCAs [23].

Another certificate revocation protocol for ad hoc networks, that provides a measure of protection against malicious accusation attacks, was proposed in [25]. Information that are used to decide whether or not a certificate should be revoked, is shared by all the nodes; however, it is the individual nodes that are given the responsibility of revoking certificates and storing information about the status of the certificates of the peers they communicate with. Prior to entering a network, a node is required to have a valid certificate issued by a CA that is trusted by the other network peers. It is also expected to have the public keys of the CAs that issued the certificates of the peers it expects to communicate with. The first duty of a node after entering a network is to broadcast its certificate to all the nodes, and simultaneously sends a request that the nodes send their profile tables. The profile table contains information about the behavior profile of each node in a network. The information in the profile tables is used to determine whether or not a given certificate should be revoked. Each node is required to compile and maintain a profile table. A profile table can be represented in the form of a packet of varied length depending on the number of accusation launched against the nodes [25].

Support for distributed node revocation: using the voting scheme; was proposed in [28]. If any node observes more than some threshold votes against some node A they break off communications with A. The base station can relay votes to a physical secure location where undeployed nodes are stored& they erase pairwise keys with A from undeployed key rings.

4 REPUTATION SCHEMES

In this section we focus on the reputation and trust schemes that have been proposed for ad hoc networks. In section 4.1, the concept, goals, features, and architecture of reputation systems are presented, whereas in section the reputation and trust based security schemes are surveyed.

4.1 System Goals Features and Architecture

In mobile ad hoc networks, nodes are both routers and terminals. For lack of routing infrastructure, they have to cooperate to communicate. Cooperation at the network layer means routing and forwarding packets. Misbehavior means [29] deviation from regular routing and
forwarding. It arises for several reasons; unintentionally when a node is faulty. There is a natural incentive for nodes [29] to only consume, but not contribute to the services of the system. Intentional misbehavior can aim at an advantage for the misbehaving node or just constitute vandalism, such as enabling a malicious node to mount an attack or a selfish node to save power. The use of reputation systems in many different areas of IT is increasing, they are used to decide who to trust, and to encourage trustworthy behavior. Resnick and Zeckhauser [30] identify three goals for reputation systems:

1. To provide information to distinguish between a trustworthy principal and an untrustworthy principal.
2. To encourage principals to act in a trustworthy manner.
3. To discourage untrustworthy principals from participating in the service the reputation mechanism is present to protect.

The features of a reputation system can be classified as follows [1 in reputation]:

Representation of information and classification: These determine how monitored events are stored and translated into reputation ratings, and how ratings are classified for response.

Use of second-hand information: Reputation systems can either rely exclusively on their own observations or also consider information obtained by others. Secondhand information can, however, be spurious, which raises the questions of how to incorporate it in a safe way and whether to propagate it.

Trust: The use of trust influences the decision of using second-hand information. The design choices are about how to build trust, out-of-hand trust versus building trust on experience, how to represent trust, and how to manage the influence of trust on responses.

Redemption and secondary response: When a node has been isolated, it can no longer be observed. The question of how those nodes should be rated over time is addressed by these two features. If the misbehavior of a node is temporary, a redemption mechanism ensures that it can come back to the network. It is, however, desirable to prevent recidivists from exploiting a redemption mechanism. This can be achieved by secondary response, meaning a quicker response to a recurring threat, in analogy to the human immune system [1 in reputation].

To enable nodes to adapt to changes in the network environment caused by misbehaving nodes, a detection & reputation system consists of three modules [1 in reputation], monitoring, reputation and response modules. The goal of monitoring is to gather first hand information about the behavior of nodes in a network. The two main ideas behind reputation that it is used as an incentive for good behavior and provides a basis for the choice of transaction partners. The response aims at isolating misbehaving nodes. This isolation has three purposes. The first is to reduce the effect of misbehavior by depriving the misbehaving node of the opportunity to participate in the network. The second is to serve as an incentive to behave well to not be denied service. Finally, the third is to obtain better service. Figure 5 summarizes the goals features and architecture of a reputation system designed for ad hoc networks.

4.2 Reputation Schemes in Ad Hoc Networks

This section we focus on the reputation and trust schemes that were suggested for ad hoc networks and give a survey of these schemes. In [31] a trust model for mobile ad hoc networks was introduced. Initially each node is assigned a trust level. Several approaches are used to dynamically update trust levels by using reports from threat detection tools, such as Intrusion Detection Systems (IDSs), located on all nodes in the network. The nodes neighboring to a node exhibiting suspicious behavior initiate trust reports. These trust reports are propagated through the network. A source node can use the trust levels it establishes for other nodes to evaluate the security of routes to destination nodes. Using these trust levels as a guide, the source node can then select a route that meets the security requirements of the message to be transmitted. Important concepts are demonstrated for establishing a collaborative, dynamic trust model and for using the proposed model as an example to enhance the security of message routing in mobile ad hoc networks.

In [32], a method to distinguish selfish peers from cooperative ones was developed based solely on local observations of AODV routing protocol behavior. The approach uses the finite state machine model of locally observed AODV actions to build up a statistical description of the behavior of each neighbor. A series of well known statistical tests to features derived from this description are applied to partition the set neighboring nodes into a cooperative and selfish class.

A node can have a reputation value about a subject without ever having interacted with it himself. However, an inherent problem with any such mechanism is the vulnerability to liars. Untrustworthy nodes can have different strategies to publish their falsified first-hand information when attempting to influence reputation ratings (e.g., when they want to discredit regular nodes). The basic strategies are changing reported misbehavior instances, reported regular behavior, both, mixed, or applied only occasionally. [29] Liars may also use the following strategies [29]:

Brain washing: When a node is surrounded by colluding lying nodes, it can be tricked into believing false information. When it later moves into a
Figure 5: Goals, features and architecture of an ad hoc networks reputation system.

Network-level security can increase due to the democratic voting mechanism of independent measurement entities, each independently aiming at a higher security level in the network. In [34], elements of a monitoring scheme in MANETs were presented. It was stated that a security monitoring system continuously estimating the actual security level can be attached to individual nodes. There are two separate goals in estimation process in [34]: security level of node and security level of network. The elements of the architecture are a measurement entity (ME) attached to each node and a voting entity (VE), trusted entity attached to a node trusted by a group of nodes with MEs. Each ME in the network maintains a private reputation repository of the network elements with the following information for each metric (metric objects, metric methods, and metric measurement rod). In addition to the metric repository of the network elements of a MANET, A VE contains the same functionality as ME, in addition, it has an organizer role in case of several MEs are going to make decisions concerning the security level & trustworthiness of a node. Certain trusted nodes can act as VEs in an AHN. A countermeasure entity CME acts on the results obtained from the voting process. A trust establishment mechanism is needed to enable estimation & voting process and to select VEs, CMEs. Figure 6 depicts the democratic voting situation, the phases are as follows:

- An ME detects suspicious activity in the different neighborhood with honest nodes, it will not believe them since their information deviates too much from its own.

Intoxication: Nodes could try to gain trust from others by telling the truth over a sustained period of time and then start lying.

Identity spoofing: Without identity persistence, a badly rated node could disappear and reappear with a different identity.

By using second hand information, an accurate estimate of some subject’s behavior can be obtained faster. A first step to the analysis of a reputation system based on a deviation test was presented in [33]. Nodes accept second hand information only if this does not differ too much from their reputation values. Direct observations are always accepted and the reputation values updated accordingly. An indirect (second hand) observation arises from interactions with peers who report about their own direct observations. Indirect observations are only accepted if the reported observation does not deviate too far from the current reputation. To keep a history of previous events, two counters are updated whenever there is a new observation, either direct or indirect. One of them tracks positive observations, and the other keeps track of negative observations. Direct observations are always accepted and counted with indirect observations have to pass a deviation test [33].

Redemption and secondary response
neighboring node.
- The ME reports the findings to its VE.
- The VE informs all its MEs.
- The MEs report their observations on the suspected node to the VE.
- The results are gathered by the VE and delivered to the CME and back to the MEs.
- The CME institutes countermeasures based on the voting results. For example, in the case of a remarkable threat, a node can be isolated from the network by invalidating its IP address.
- The MEs’ trust level concerning the suspected node can be updated based on the voting results and the decision making about this is left to each ME.

Figure 6: Democratic voting scheme [34]

Another collaborative mechanism for detecting malicious incorrect packet forwarding attacks was described in [35]. The proposed model provides two main functionalities: monitoring the behavior of the neighboring nodes in the network and computing their reputations based on the information provided by the monitoring. In the described trust manager protocol collaboration between neighboring nodes is required. Mechanism builds trust through the trust manager. As it is shown in Figure 7, there are two main modules the monitoring module and the reputation handling module. In the monitoring module, each node independently monitors its neighboring nodes forwarding activity. Monitoring is related to the proportion of correctly forwarded packets during a fixed time window. If anomaly is detected, monitor informs the reputation manager. The reputation handling module consists of four components, the first is the reputation collecting through sensing or direct monitoring or recommendations& accusations using on demand technique or proactive broadcasting technique. The mechanism uses proactive & on demand techniques. The second component is the reputation formatting which uses a reputation template containing different fields. The reputation information has to be evaluated before it is locally stored or broadcasted to the neighborhood. That is why in the reputation information maintenance, each node is assumed to maintain a reputation table for storing its one hop neighborhood reputation information that it gets by direct monitoring or through broadcast from some neighboring nodes. In the reputation rating module, the most recent reputation is always considered heavier.

Figure 7: Trust Manger Architecture [35]

The performance of three trust-based reactive routing protocols in a network with varying number of malicious nodes was evaluated in [36]. Every time a node transmits a data or control packet, it immediately brings its receiver into the promiscuous mode so as to overhear its immediate neighbor forwarding the packet. Two categories could be derived to compute direct trust: the first category is acknowledgment, provides with information concerning backhole, modification, attacks and the second category is packet precision for data integrity. The trusted update interval has been proved to be a very critical component, it determines the time a node should wait before assigning a trust level. In [36], each trust category is represented by one or more types of events. The successful and failed events of all categories are represented in tables, and all events are then normalized to produce usable information having statistical properties. The normalized value of one of the events used in the computation of a category is calculated a function of a failed and successful events. Trust values from the two trust categories are the assigned weights according to their priorities in order to determine the direct trust level of a particular node.

A scheme for evaluating trust evidence in ad hoc networks was presented in [37]. It is entirely based on information originating at the users of the network. No centralized infrastructure is required, although the presence of one can certainly be utilized. Also, users need not have personal, direct experience with every other user in the network in order to compute an opinion about them. They can base their opinion on secondhand evidence
A scheme that allows trust management to be performed locally on the individual ad hoc network was proposed in [12]. The first duty of a node entering in the network is to broadcast its certificate to all nodes and simultaneously send a request that the nodes send their profile tables and compiles its own profile table. Profile tables have the following fields: owner’s ID, peer’s ID, accusation information, and certificate status. In addition, status tables are used to ascertain the status of a certificate. It consists of: number of accusations against the nodes, behavior index of a node, weight of node accusation, revocation quotient, and certification status. When a certificate is revoked, all previously established trust relations for the node in question, is immediately negated by all nodes and network access consequently denied. In [39], a secure random reporting protocol for a civilian ad hoc network was proposed. In this protocol, the source and destination collect reports from intermediate nodes on the routing path. Every data packet initiates a report from one intermediate node that is randomly chosen by a source node. Through a symmetric cryptographic construction, the node selection is not disclosed to other intermediate nodes. The random reporting protocol has three modes: the basic periodic reporting, the random reporting node selection, the random reporting node and direction selection, and the random bidirectional selection. Although the report is securely transmitted to the destination, it is not guaranteed to be accurate, since nodes may cheat in order to get credit. A chained scheme has been devised on the link layer acknowledgments to verify the validity of the received report. From both security and performance perspectives, the secure random reporting protocol is advantageous for gathering the forwarding activities of mobile nodes in civilian ad hoc networks. The report can be used for determining whether congestion exists in network, engineering the traffic, crediting nodes with how many packet they relayed, and detecting that nodes maliciously drop packets.

5 AUTHENTICATION

Due to the ad hoc networks characteristics, the authentication protocols used for routing and data packet delivery in ad hoc networks should be lightweight and scalable. Asymmetric cryptography does not adapt well to ad hoc networks in that the processing required for asymmetric cryptography is very CPU intensive and the technique has been proved to be prohibitively insufficient in wireless ad hoc networks in terms of message overhead and computation complexity. Symmetric cryptography algorithms are fast. Nevertheless, they introduce complexity in key maintenance and exert difficulty in authentication for multicast or broadcast communications. Moreover, radio channels in wireless networks are more erroneous and lossy than the communication links in the Internet. With multiple receivers, there could be a high variance
among the bandwidth and radio interference of different receivers, with high packet loss for the receivers with low bandwidth and high radio interference. Threshold cryptographic solutions may not be suitable for most commercial ad hoc networks environments, for the following reasons [25]:

1. Computationally exhaustive: Threshold cryptography involves additional computationally intensive modular exponentiations compared to the underlined asymmetric-key cryptographic protocols. Most low-powered wireless nodes do not have the resources to handle such computationally intensive operations. For nodes with less resources constraints, the increase in latency due to the extra computational cost may not be acceptable. For example, the analysis of the implementation in [40] indicates that generation of a partial RSA signature using one of shares is approximately 2.5 times slower than standard RSA signing. Considering that partial signatures need to be generated then combined to obtain a valid signature, the increase in latency due to the additional computations may not be acceptable.

2. Requires unselfish cooperation: Network security solutions involving threshold cryptography require unselfish cooperation of the communicating peers. This might not be an issue in certain military applications; however, in most commercial network applications nodes may not behave unselfishly. Wireless nodes are often limited in battery power and utilize power conservation mechanisms that encourage them to remain dormant unless they are performing necessary services. It might not be realistic therefore to expect nodes in certain environments to behave unselfishly and cooperate, for example to service certificate requests.

Considering of the above problems, the authentication mechanism is expected to be effective even in the presence of high packet loss [41].

To verify the correctness of a received packet, the method to put the e-signature on the packet by the public key is basic on an ad hoc network. However, since a portable terminal used in ad hoc networks has relatively small calculation ability and a lot of calculation time is needed for giving and verification of e-signature. In [42], two methods were proposed to authenticate a consecutive packet efficiently by using a digital signature and a comparatively high-speed hash function.

A lightweight authentication protocol that effectively and efficiently provides security properties such as authenticity and integrity for communicating neighbor nodes in MANETs was proposed in [41]. The protocol utilizes one-way hash chains to compute authentication keys, which not only eliminates the high performance overhead imposed by asymmetric cryptography (such as digital signatures), but also avoids the difficulty of key management introduced by secret paired symmetric key. The protocol also used delayed key disclosure to prevent a malicious entity from forging packets with MACs with an already released key. The authentication protocol is lightweight, scalable and tolerant of packet loss. The performance analysis showed that the protocol incurs low overhead penalty and also achieves a tradeoff between security and performance.

An interleaved message authentication scheme was proposed and evaluated in [43]. Interleaved authentication is used to restrain malicious nodes from manipulating messages by implicitly monitoring their actions. A node must share keys with all nodes within a radius of k-hops. A receiving node expects k authentication codes from different nodes in order to accept a message, if at least one of them does not match the message content, the message is rejected. This means that sets up to k-1 collaborating malicious nodes are prevented. Figure 9 depicts a communication path with interleaved message authentication with k=2.

![Figure 9: A communication path with interleaved message authentication (k=2) [43]](image-url)

Also shortcuts are used for authentication; shortcuts are links that are established between distant nodes. Each node stores a certain, small, number of keys that re-enforce the path from a message source to its destination. When a message is sent, it is first routed to the shortcut node that is closest to the message target. On its way from the shortcut node to the destination, the message is authenticated with the basic Canvas protocol. Interleaved paths from shortcuts are also built to span very large distances.

A solution that accomplishes end-to-end authentication of ACKs based on the TESLA symmetric key broadcast authentication protocol was proposed in [44]. The scheme provides a dependable and inexpensive solution to rating packet forwarding services in clustered ad hoc networks with centralized supervision. Authentication performance is based on two factors: threshold level and authentication delay. In [38] the authentication delay was considered. While a centralized architecture can guarantee the authentication delay, this is not possible in a distributed authentication scheme where nodes are mobile. Security impact on QoS in a distributed system was investigated by looking at local and global schemes for achieving security while maximizing QoS. An intelligent approach to determine the optimum threshold level (OTL) under
different conditions was proposed.

An anonymous on-demand routing protocol, termed MASK, to enable anonymous communications thereby thwarting possible traffic analysis attacks was proposed in [45]. Based on a new cryptographic concept called pairing, an anonymous neighborhood authentication protocol which allows neighboring nodes to authenticate each other without revealing their identities was suggested. The secret pairwise link identifiers and keys established between neighbors were utilized during the neighborhood authentication process. MASK fulfills the routing and packet forwarding tasks nicely without disclosing the identities of participating nodes under a rather strong adversarial model. It also provides the desirable sender and receiver anonymity, as well as the relationship anonymity of the sender and receiver. It is also resistant to a wide range of adversarial attacks; moreover, it preserves the routing efficiency in contrast to previous proposals.

A protocol called SDF which provides a solution for secure data forwarding in wireless ad hoc networks was presented in [46]. The protocol can detect and locate faulty links on a per packet basis so that an appropriate action can be taken. It provides authentication using efficient hash chains and one-time hash tag commitments. The simulation results show that the SDF-enhanced AODV is as efficient as the plain AODV in discovering and maintaining routes for delivery of data packets, at the cost of using larger routing packets and adding data control packets which result in a higher overall bytes overhead, and in exchange for a slightly higher packet delivery latency because of the cryptographic computation incurred.

An Authentication Service Based on Trust and Clustering in Wireless Ad Hoc Networks was described and evaluated in [18]. It is a combined reputation and authentication scheme in which there are two types of trust: direct within same cluster and recommended between different clusters. For certification within the same cluster, there is no problem as nodes know each other. For certification within different groups, the node selects n nodes (called introducers) with the highest trust values and sends them request messages. Before sending out the request message, node v_i first checks whether it is in the same cluster as v_j. If it is, it sends the request message to its neighboring nodes, assuring that some of its neighboring nodes have built up a direct trust relationship with v_j. On the other hand, if v_i and v_j are in different clusters, then the problem becomes more complicated. Node v_i has to select some trustworthy nodes in the target cluster to be the introducing nodes, or so-called introducers, they are nodes in the same cluster as v_j for which v_i has high trust values. However, it is possible for the introducers to be malicious; therefore, a voting procedure is carried out to conclude the correct public key of the target node by majority vote. Identification and isolation of malicious nodes is done using three methods. First Method: Direct monitoring of individual nodes by listening to the traffic via wireless communications using a monitoring facility. Second: By identifying suspicious introducers who provide public key certificates different from the others. Third: If the trust values provided by the introducer indicate a node is malicious. To deal with colluding nodes a scheme is suggested. After filtering out suspicious introducers, the trust value of a target node t is obtained from the rest of introducers.

The concept of how nodes should be identified and authenticated was addressed in [47]. After discussing related works and the concept of identities and identifiers in MANETs, the MANET-ID system, which can be used to reliably identify nodes in an ad hoc network with properties like uniqueness, irreversible ties with the identified object, immutability throughout the lifetime of the object and non-transferability, was presented.

6 CONCLUSIONS AND CHALLENGES

In this paper we surveyed some of the security approaches used for securing ad hoc networks. These are approaches the threshold cryptography, certification authorities, reputation and trust, and authentication. There are still many challenges and research openings in the area of ad hoc networks security. Although there were suggestions for the optimum threshold level for threshold cryptography, however there is still a need for more research to answer many questions as: What are the upper and lower bound threshold values and the optimum threshold value. Also, is the partial key provided valid all the time? What if corrupted nodes provide incorrect partial keys? Can error correcting codes be used in conjunction with threshold cryptography to compensate for the effects of malicious partial key shares? What about the dynamic adjustment of the partial key validity time? Also more research is needed to compare between fixed and dynamic threshold levels, taking into consideration the geographical distribution of nodes in ad hoc networks.

Also despite the great effort that has been consumed in the study and design of certificate distribution schemes, there are still lots of openings and challenges in this area. For example there is no clear criteria for the CAs selection such as depending on their roles, power, reputation, age in the network, etc. Also the number of CAs with respect to the total number of nodes in the network, and their distribution needs to be formulated while taking into considerations the network topology and the mobility of the nodes within the network which dynamically affects the nodes’ distribution within
the network. Some schemes have suggested a time out for certificates, it needs to be calculated as well. For CAs revocation voting and reputation schemes can be used to gain a better judgment on a CA behavior, to isolate it and discard certificates issued from that CA. Moreover, a lightweight method for propagating the revocation news needs to be investigated to decide whether the periodic announcement or the on demand is more suitable in the case of ad hoc networks. We surveyed the different reputation and trust based schemes that were proposed for ad hoc networks in the literature ranging between collaborative and independent node based schemes. Several reputation schemes can be modified or blended together to enhance their performance and obtain an optimum scheme that is suitable to the ad hoc networks very specific characteristics. For example the secure random reporting protocol that was proposed in [12in reputation] can be modified by assigning different weights to the nodes’ reports according to the reputation of the node issuing the report.

Some of the authentication schemes proposed in the literature need to be combined with other security schemes like reputation and trust based schemes. In the future we plan to investigate some of those challenging research areas such to obtain a more secure scheme for ad hoc networks.

In the future we plan to investigate some of those challenging research areas such as to obtain a more secure scheme for ad hoc networks.

REFERENCES


