Towards reliable network services in Ad Hoc Networks:

Protecting the routing protocols

Anne Marie Hegland

PhD Dissertation
Department of Informatics,
Faculty of Mathematics and Natural Sciences,
University of Oslo
To my son Andreas and my husband Tom Erik
Abstract

With the wireless media and ad hoc nodes operating both as routers and communication end-terminals, the ad hoc routing protocols are more prone to attacks than the network layer of fixed networks. Authenticated routing messages are a primary concern in the provisioning of a reliable network service.

The scope of this dissertation has been network layer security in ad hoc networks intended for the operational scenario of emergency and rescue operations. The main objectives have been to investigate the applicability of existing security schemes for ad hoc networks, and to contribute to the development of resource efficient, yet robust and appropriate protection schemes.

The majority of secure ad hoc routing protocols proposed so far, tend to focus on the protection techniques, rather than the computational cost and bandwidth consumption. Our work highlights the importance of taking these factors into consideration in the design of secure routing protocols. In addition, attention must be paid to the nature of routing control traffic and its discrepancies from ordinary application data.

The thesis is based on our research published in four articles. Our contributions include a comprehensive survey of key management methods proposed for ad hoc networks and an evaluation of the applicability for initialization and maintenance of a protected network service. The applicability of identity-based public key schemes for protection of ad hoc routing information is also analyzed. Scalability remains an issue even with ID-based schemes that remove the need for bandwidth consuming certificate exchanges. A hybrid protection scheme is therefore proposed for the Optimized Link-State Routing protocol (OLSR). The protection scheme reduces both computational cost and bandwidth consumption by combining identity-based signatures with values from a hash chain. And last, a simple and robust, yet scalable, method is introduced for the distribution of revocation information in ad hoc networks.
Preface

This dissertation is submitted to the Department of Informatics, Faculty of Mathematics and Natural Sciences, University of Oslo, in fulfillment for the degree of Philosophiae Doctor (PhD). The research was carried out at the University Graduate Centre (UniK) at Kjeller under the supervision of Professor Emeritus Pål Spilling and Associate Professor Leif Nilsen from the University of Oslo, and the co-supervision of Professor Øivind Kure from the Norwegian University of Science and Technology, Trondheim.

My PhD is part of the SIMULTAN (Secure IP-based MULti media Access Network) program at Kongsberg Defence & Aerospace. The research has been funded by the Norwegian Research Council (grant no. 156713/140 “Ad-hoc trådløse nettverk. Sikkerhet i trådløse nettverk”) and Kongsberg Defence & Aerospace. The PhD is also included in the Future Communications Systems (FUCS) program at UniK.
Acknowledgements

Fist of all I would like to express my thanks to my supervisors, Professor Emeritus Pål Spilling, Professor Øivind Kure and Associate Professor Leif Nilsen, for their advices, encouragement and constructive criticism. I also thank Mona Holsve Ofigsbø and my colleagues from the ad hoc research group at UniK for all good discussions. A special thank to Eli Winjum, Professor Stig Frode Mjølsnes and Professor Chunming Rong for co-authoring a couple of my articles. In addition, I owe a great thank to all the anonymous reviewers for their helpful comments and suggestions. I am also grateful to Professor Loren Olsson at the University of Tromsø who helped proof-reading one of my manuscripts.

I thank my employer, Kongsberg Defence & Aerospace, for giving me the opportunity and financial support to carry out the PhD project. Thanks to my colleagues at the Security and Concept Development Department for discussions. A special thank to Ragnar Wik for inspiration and back up. Furthermore, I would like to express my thanks to the service minded staff at UniK.

Last, but not least, I thank my family for their support. Thank you for your patience, Tom and Andreas. And thanks to my mother Aud, my stepdaughter Linda and her dear Rasmus for baby-sitting.
LIST OF PUBLICATIONS

Papers I through IV represent the main contributions of this thesis. The articles are found in Part II. Paper V – VIII are related work. I am the first author of papers I through VI, and a co-author of papers VII and VIII. Whereas all papers were written in co-operation with other authors, I conducted the core of the work in papers I through VI. My contributions to papers VII and VIII were as a discussion partner.

Main contributions:


Related Work:


# Table of Contents

**Abstract**................................................................................................................................V

**Preface**.....................................................................................................................................VII

**Acknowledgements** ................................................................................................................IX

**List of Publications** ................................................................................................................XI

## Part I .................................................................................................................................1

**Abbreviations**.....................................................................................................................3

1 **Introduction** ......................................................................................................................6

1.1 **Background and Motivation** ........................................................................................6

1.2 **Scope and Objectives** ..................................................................................................6

1.3 **Structure of the Thesis** ...............................................................................................8

1.4 **Methods** ......................................................................................................................8

1.5 **Setting the Scene: Mobile Ad Hoc Networks for Emergency and Rescue Operations** ....9

1.5.1 Organizational structure of emergency and rescue operations ................................. 9

1.5.2 Characteristics of MANETs for emergency and rescue operations ....................... 10

1.5.3 Network architecture .................................................................................................. 10

1.5.4 A note on MANETs and TETRA ............................................................................ 12

1.5.5 Summary .................................................................................................................... 12

1.6 **Threats to Ad Hoc Networks** ....................................................................................13

1.6.1 Threats: MANETs versus infrastructure based wireless networks ...................... 13

1.6.2 Threats to MANET routing protocols .................................................................. 13

1.6.3 Threat model ............................................................................................................ 15

1.6.4 Summary .................................................................................................................... 16

1.7 **A Summary of the Thesis** ...........................................................................................16

2 **Security in Ad Hoc Networks** .......................................................................................19

2.1 **Ad Hoc Routing Protocols** ........................................................................................19

2.1.1 Classification of ad hoc routing protocols ................................................................19

2.1.2 Standardization work .................................................................................................20

2.1.3 Proactive ad hoc routing protocols ..........................................................................21

2.1.3.1 OLSR and OLSRv2 .......................................................................................... 21

2.1.3.2 OSPF-MANET protocols .............................................................................. 22

2.1.4 Reactive ad hoc routing protocols ...........................................................................23

2.1.4.1 DSR ................................................................................................................ 23

2.1.4.2 AODV and DYMO ....................................................................................... 23

2.2 **Protecting the Ad Hoc Routing Protocols** ................................................................24

2.2.1 Trust model ..............................................................................................................24

2.2.2 Definition of security for ad hoc routing protocols ................................................ 25
2.2.3 Basic protection mechanisms ................................................................. 26
  2.2.3.1 Message authentication codes ........................................................... 26
  2.2.3.2 Cryptographic signatures ................................................................. 27
  2.2.3.3 Hash Chains .................................................................................... 28
  2.2.3.4 Redundancy .................................................................................... 29
  2.2.3.5 Reputation schemes ......................................................................... 29
  2.2.3.6 Message identifiers .......................................................................... 30
  2.2.3.7 Packet leashes ................................................................................. 30
  2.2.3.8 Use of the protection mechanisms .................................................... 31
2.2.4 Basic security modules ........................................................................... 33
  2.2.4.1 TESLA ............................................................................................. 33
  2.2.4.2 CONFIDANT .................................................................................. 34
  2.2.4.3 Watchdog and Pathrater ................................................................. 34
  2.2.4.4 TIARA ............................................................................................ 35
  2.2.4.5 SAR ............................................................................................... 35
  2.2.4.6 TMR ............................................................................................... 36
2.2.5 Secure ad hoc routing protocols .............................................................. 36
  2.2.5.1 Secure proactive ad hoc routing protocols ......................................... 36
    2.2.5.1.1 SEAD ....................................................................................... 36
    2.2.5.1.2 SLSP ........................................................................................ 37
    2.2.5.1.3 Security extensions to OLSR ...................................................... 37
  2.2.5.2 Secure reactive ad hoc routing protocols ............................................ 38
    2.2.5.2.1 SAODV .................................................................................... 38
    2.2.5.2.2 Ariadne ....................................................................................... 39
    2.2.5.2.3 SRP .......................................................................................... 40
    2.2.5.2.4 ARAN ....................................................................................... 40
    2.2.5.2.5 SPAAR ..................................................................................... 41
2.2.6 Security extensions proposed in the MANET WG and OSPF WG Internet-
Drafts and RFCs ............................................................................................. 42
  2.2.7 On the applicability of IPsec and IEEE 802.11i for ad hoc networks ...... 42
  2.2.8 Summary ............................................................................................. 44
2.3 Key Management ....................................................................................... 44
  2.3.1 Key generation and authentication ......................................................... 44
  2.3.2 Key Lengths and security levels ............................................................ 45
  2.3.3 Revocation and exclusion of nodes ....................................................... 47
  2.3.4 Summary ............................................................................................. 48
3 Contributions of the Thesis .......................................................................... 50
  3.1 Contribution of Paper I: A Survey of Key Management in Ad Hoc
Networks ....................................................................................................... 51
  3.2 Contribution of Paper II: Analysis of IBS for MANET Security in
Emergency and Rescue Operations ................................................................ 52
  3.3 Contribution of Paper III: Hybrid Protection of OLSR ......................... 53
  3.4 Contribution of Paper IV: On the Distribution of Revocation
Information in Ad Hoc Networks ................................................................. 55
4 Concluding Remarks .................................................................................... 57
References ...................................................................................................... 58
PART II...........................................................................................................67

PAPER I: A SURVEY OF KEY MANAGEMENT IN AD HOC NETWORKS....69
  I INTRODUCTION .......................................................................................... 72
  II DESIRABLE FEATURES OF MANET KEY MANAGEMENT SCHEMES .......... 74
  III OVERVIEW OF KEY MANAGEMENT SCHEMES ........................................... 76
  IV CONCLUSIONS ............................................................................................ 107
  ACKNOWLEDGEMENTS .................................................................................... 109
  REFERENCES .................................................................................................... 109

PAPER II: ANALYSIS OF IBS FOR MANET SECURITY IN EMERGENCY AND RESCUE OPERATIONS .................................................................119
  1 INTRODUCTION ................................................................................................ 121
  2 EMERGENCY AND RESCUE OPERATIONS ..................................................... 122
  3 IDENTITY-BASED SIGNATURE SCHEMES ...................................................... 122
  4 THE OPTIMIZED LINK STATE ROUTING PROTOCOL (OLSR) ...................... 123
  5 PROTECTING MANET ROUTING PROTOCOLS WITH IBS ............................ 124
  6 RELATED WORK ............................................................................................. 129
  7 CONCLUSIONS AND FUTURE WORK ........................................................... 129
  REFERENCES .................................................................................................... 130

PAPER III: HYBRID PROTECTION OF OLSR......................................................... 133
  1 INTRODUCTION ................................................................................................ 135
  2 THE OPTIMIZED LINK STATE ROUTING PROTOCOL (OLSR) ...................... 137
  3 THREATS .......................................................................................................... 138
  4 BASIC SECURITY MECHANISMS .................................................................. 139
  6 PERFORMANCE ............................................................................................... 145
  7 RELATED WORK ............................................................................................. 149
  8 CONCLUSIONS AND FUTURE WORK ........................................................... 150
  REFERENCES .................................................................................................... 151
  A APPENDIX: SECURITY LEVELS, HASH LENGTHS AND SIGNATURE SIZES ...... 153

PAPER IV: ON THE DISTRIBUTION OF REVOCATION INFORMATION IN AD HOC NETWORKS........................................................................ 155
  I INTRODUCTION ................................................................................................ 157
  II RELATED WORK ............................................................................................ 159
  III REQUIREMENTS ........................................................................................... 160
  IV THE MRL SCHEME ......................................................................................... 161
  V THE MRL SCHEME WITH VARIOUS ROUTING PROTOCOLS .................... 164
  VI ANALYSIS ..................................................................................................... 168
  VII CONCLUSIONS AND FUTURE WORK ......................................................... 174
  REFERENCES .................................................................................................... 174
LIST OF FIGURES

Figure 1 Organizational structure of emergency and rescue operations......................... 9
Figure 2 Network architecture .......................................................................................... 11
Figure 3 Various options for protection of ad hoc routing information ......................... 43
Figure 4 Outline of the work and contributions.............................................................. 50

LIST OF TABLES

Table 1 Features of symmetric and asymmetric schemes for protection of ad hoc routing information.............................................................................................................. 31
Table 2 Comparable security levels................................................................................ 46
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AES</td>
<td>Advanced Encryption Standard</td>
</tr>
<tr>
<td>AH</td>
<td>Authentication Header</td>
</tr>
<tr>
<td>ANSI</td>
<td>American National Standards Institute</td>
</tr>
<tr>
<td>AODV</td>
<td>Ad hoc On-demand Distance Vector routing protocol</td>
</tr>
<tr>
<td>ARAN</td>
<td>A secure Routing protocol for Ad hoc Networks</td>
</tr>
<tr>
<td>CA</td>
<td>Certificate Authority</td>
</tr>
<tr>
<td>CAM</td>
<td>Child-proof Authentication for MIPv6</td>
</tr>
<tr>
<td>CBC</td>
<td>Cipher-Block-Chaining</td>
</tr>
<tr>
<td>CONFIDANT</td>
<td>Co-operation Of Nodes: Fairness In Dynamic Ad-hoc NeTworks</td>
</tr>
<tr>
<td>CRL</td>
<td>Certificate Revocation List</td>
</tr>
<tr>
<td>CRS</td>
<td>Certificate Revocation Status</td>
</tr>
<tr>
<td>DD</td>
<td>Database Description</td>
</tr>
<tr>
<td>DES</td>
<td>Data Encryption Standard</td>
</tr>
<tr>
<td>DMO</td>
<td>Direct Mode Operation</td>
</tr>
<tr>
<td>DoS</td>
<td>Denial of Service</td>
</tr>
<tr>
<td>DSDV</td>
<td>Destination-Sequenced Distance-Vector</td>
</tr>
<tr>
<td>ESP</td>
<td>Encapsulating Security Payload</td>
</tr>
<tr>
<td>ETSI</td>
<td>European Telecommunications Standards Institute</td>
</tr>
<tr>
<td>FRAC</td>
<td>Flow-based Route Access Control</td>
</tr>
<tr>
<td>GMAC</td>
<td>Galois Message Authentication Codes</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Positioning System</td>
</tr>
<tr>
<td>HMAC</td>
<td>keyed-Hash Message Authentication Code</td>
</tr>
<tr>
<td>HNA</td>
<td>Host and Network Association</td>
</tr>
<tr>
<td>IBS</td>
<td>Identity-Based Signature</td>
</tr>
<tr>
<td>ID</td>
<td>Identifier</td>
</tr>
<tr>
<td>IDS</td>
<td>Intrusion Detection System</td>
</tr>
<tr>
<td>IEEE</td>
<td>International organization for Electrical and Electronics Engineers</td>
</tr>
<tr>
<td>IETF</td>
<td>Internet Engineering Task Force</td>
</tr>
<tr>
<td>IP</td>
<td>Internet Protocol</td>
</tr>
<tr>
<td>KAC</td>
<td>Key Administration Centre</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Kbps</td>
<td>kilobits per second</td>
</tr>
<tr>
<td>LLS</td>
<td>Link-Local Signaling</td>
</tr>
<tr>
<td>LSU</td>
<td>Link State Update</td>
</tr>
<tr>
<td>MAC</td>
<td>Media Access Control or Message Authentication Code</td>
</tr>
<tr>
<td>MANET</td>
<td>Mobile Ad hoc NETwork</td>
</tr>
<tr>
<td>MID</td>
<td>Multiple Interface Declaration</td>
</tr>
<tr>
<td>MDR</td>
<td>MANET Designated Router</td>
</tr>
<tr>
<td>MPR</td>
<td>MultiPoint Relay node</td>
</tr>
<tr>
<td>MRL</td>
<td>MANET Revocation List</td>
</tr>
<tr>
<td>NIST</td>
<td>National Institute of Standards and Technology</td>
</tr>
<tr>
<td>OCSP</td>
<td>Online Certificate Status Protocol</td>
</tr>
<tr>
<td>OLSR</td>
<td>Optimized Link-State Routing protocol</td>
</tr>
<tr>
<td>OSPF</td>
<td>Open Shortest Path First</td>
</tr>
<tr>
<td>PGP</td>
<td>Pretty Good Privacy</td>
</tr>
<tr>
<td>PhD</td>
<td>Philosophiae Doctor</td>
</tr>
<tr>
<td>RERR</td>
<td>Route ERRor</td>
</tr>
<tr>
<td>RFC</td>
<td>Request For Comments</td>
</tr>
<tr>
<td>RM</td>
<td>Routing Message</td>
</tr>
<tr>
<td>RREP</td>
<td>Route REPply</td>
</tr>
<tr>
<td>RREQ</td>
<td>Route Request</td>
</tr>
<tr>
<td>SAR</td>
<td>Security-aware Ad hoc Routing</td>
</tr>
<tr>
<td>SAODV</td>
<td>Secure Ad-hoc On-demand Distance Vector</td>
</tr>
<tr>
<td></td>
<td>or Security-aware AODV (depending on context)</td>
</tr>
<tr>
<td>SEAD</td>
<td>Secure Efficient distance vector routing protocol for AD hoc networks</td>
</tr>
<tr>
<td>SIMULTAN</td>
<td>Secure IP-based MULTI-media Access Network</td>
</tr>
<tr>
<td>SLSIP</td>
<td>Secure Link State routing Protocol</td>
</tr>
<tr>
<td>SMF</td>
<td>Simple Multicast Forwarding</td>
</tr>
<tr>
<td>SRP</td>
<td>Secure Routing protocol</td>
</tr>
<tr>
<td>SUCV</td>
<td>Statistically Unique and Cryptographically Verifiable</td>
</tr>
<tr>
<td>TBRPF</td>
<td>Topology Dissemination Based on Reverse-Path Forwarding</td>
</tr>
<tr>
<td>TC</td>
<td>Topology Control</td>
</tr>
<tr>
<td>TDMA</td>
<td>Time Division Multiple Access</td>
</tr>
<tr>
<td>TESLA</td>
<td>Timed Efficient Stream Loss-tolerant Authentication</td>
</tr>
<tr>
<td>TETRA</td>
<td>TErrestrial Trunked RAdio</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>TIARA</td>
<td>Techniques for Intrusion Resistant Ad hoc Routing Algorithms</td>
</tr>
<tr>
<td>TIK</td>
<td>TESLA with Instant Key disclosure</td>
</tr>
<tr>
<td>TLV</td>
<td>Type-Length-Value</td>
</tr>
<tr>
<td>TMO</td>
<td>Trunked Mode Operation</td>
</tr>
<tr>
<td>TTL</td>
<td>Time To Live</td>
</tr>
<tr>
<td>TTP</td>
<td>Trusted Third Party</td>
</tr>
<tr>
<td>UDP</td>
<td>User Datagram Protocol</td>
</tr>
<tr>
<td>UniK</td>
<td>Universitetsstudiene på Kjeller</td>
</tr>
<tr>
<td>VANET</td>
<td>Vehicular Ad hoc NETwork</td>
</tr>
<tr>
<td>W-OSPF</td>
<td>Wireless OSPF</td>
</tr>
<tr>
<td>WG</td>
<td>Working Group</td>
</tr>
<tr>
<td>WSN</td>
<td>Wireless Sensor Network</td>
</tr>
<tr>
<td>ZRP</td>
<td>Zone Routing Protocol</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 Background and motivation

Wireless communication is the key to network availability anywhere and at anytime. Today’s wireless communication systems usually depend on pre-established communication infrastructure. Mobile ad hoc networks (MANETs) allow communication where no such infrastructure is available. They are self-organizing and self-configuring. The nodes operate both as communication end-points and routers. Multi-hop communication is enabled through co-operating nodes and ad hoc routing protocols.

The application areas of MANETs range from conference hall networks to military tactical operations and disaster relief. This dissertation emphasizes the latter. The PhD project is part of the SIMULTAN (Secure IP-based multi-media access networks) program at Kongsberg Defence & Aerospace. Communication solutions for emergency and rescue operations are the focal point of the program.

MANETs are expected to play an important role in future communication systems for disaster relief operations where one cannot rely on a fixed infrastructure. Ad hoc networks can also be used to extend the coverage area or reduce the load on base stations and access points in infrastructure based communication systems.

The users can be expected to anticipate the same set of services as they are used to have in the wired network. However, to meet these objectives, wireless multi-hop communication faces additional challenges and constraints compared to wired networks. Problems, ranging from power consumption and frequency allocation to routing and security, are topics for research.

1.2 Scope and Objectives

Attacks can be launched towards any layer of the protocol stack. Our scope has been network layer security in mobile ad hoc networks. The aim is a reliable network service in emergency and rescue operations.

Zhou and Haas [113] suggest redundancy (multiple routes) and cryptographic schemes for protection of ad hoc routing protocols. We have focused on cryptographic
methods. Key management is a basic problem for any scheme relying on cryptographic methods, and an initial objective was to provide an overview of proposed key management schemes and evaluate their applicability for the deployment of a protected network service.

The wireless environment and battery powered nodes make bandwidth efficiency and energy saving important design criterions for ad hoc networks. Besides limited energy and spectrum resources, the ability to scale is an important factor.

There may be a large number of recipients of ad hoc routing messages. A separate message authentication code for each possible recipient of a routing message represents a scalability problem. The receivers may not even be known in advance. Hence, pairwise symmetric keys is no viable solution. A symmetric group key scales better. The price is lower intrusion tolerance. That is, a single captured node means the system security is compromised. Furthermore, the exclusion of compromised nodes necessitates re-keying. Key changes have negative impacts on network availability during the transient period. The robustness is enhanced with asymmetric keys and exclusions through a public key revocation scheme. However, asymmetric cryptographic schemes are known to be computationally expensive and introduce more overhead than symmetric schemes.

Identity-based schemes [95] make certificate exchanges obsolete. This implies lower bandwidth cost than traditional public key schemes. The scope therefore included an investigation of the limitations of asymmetric schemes, and identity-based schemes in particular, for the protection of ad hoc routing information. A natural extension has been to explore how hybrid protection schemes using multiple security mechanisms, can be combined with the exploitation of the nature of the routing protocols to provision a reliable network service.

Assuming a protected network service with routes set up through trusted nodes only, the exclusion of compromised and malfunctioning nodes can be achieved with the aid of a revocation scheme. Revocation schemes designed for the fixed net typically assume guaranteed network connectivity and/or no bandwidth constraints. The last objective was development of a method for efficient and appropriate distribution of revocation information in ad hoc networks.
1.3 Structure of the thesis

The thesis is divided into two parts. Part I (this part) gives the motivation for the problems investigated and an introduction to the research field. Knowledge of ad hoc routing protocols is required in order to be able to suggest appropriate protection mechanisms. A survey of some representative ad hoc routing protocols is therefore included. Part I in addition discusses basic protection mechanisms and their use in the securing of route control traffic. Our contributions and some remaining problems are also described in Part I. Part II contains the publications that constitute the major results of the research conducted:

- Paper I: *Survey of Key Management in Ad Hoc Networks*
- Paper II: *Analysis of IBS for MANET Security in Emergency and Rescue Operations*
- Paper III: *Hybrid Protection of OLSR*
- Paper IV: *On the Distribution of Revocation Information in Ad Hoc Networks*

The reference section of Part I, lists the publications cited in Part I only. Publications cited in the articles in Part II are found in the reference section of each paper.

1.4 Methods

The survey paper (paper I) is a result of an extensive literature study. A number of evaluation criteria were defined and the surveyed schemes were evaluated in accordance with these criteria. For the remaining work, the problems were identified, requirements for beneficial solutions were defined, related work was studied and new solutions proposed. The suggested solutions were then evaluated and unsolved problems identified. The analytical proof-of-concepts in paper II and paper III are based on simulation results from [106], obtained with the ns-2 network simulator.
1.5 Setting the scene: mobile ad hoc networks for emergency and rescue operations

1.5.1 Organizational structure of emergency and rescue operations

Emergency and rescue operations are governed by international conventions and national laws and regulations. The organizational structure of the emergency and rescue services in Norway is specified in [53].

The emergency and rescue team on the site of the incident consists of professional units. Typically, health personnel, fire brigade and police take part. Military forces and humanitarian organizations may also be included. Figure 1 illustrates the generic organizational structure.

The emergency and rescue team reports to the rescue management on the site of the accident. The operation is coordinated from one or more off-site rescue coordination centres. In Norway, depending on the severity and size of the incident, the operation may be administered from a local rescue coordination centre or one of the two national rescue coordination centers.

![Figure 1 Organizational structure of emergency and rescue operations](image)

Altogether, whereas the time and location of the next operation are not known in advance, the involved parties and roles are to a large extent predefined. The organizational structure enables pre-configuration and pre-distribution of credentials such as cryptographic keys. The rescue management and co-ordination centers represent common points of trust. These can be exploited in the generation and distribution of keys. The key material can be distributed in the preparation phase prior to the actual operations.
1.5.2 Characteristics of MANETs for emergency and rescue operations

MANETs for emergency and rescue operations will have a planned origin. The network size and range will depend on the type of operation. Small to medium size and range will expectedly dominate. Node mobility may vary. The on-site rescue management is normally vehicle mounted. Nodes with heterogeneous capabilities can be expected. Nodes carried by the ordinary rescue personnel will be battery powered. Vehicle mounted nodes may be less energy constrained and have better physical protection. The wireless media implies limited bandwidth.

A communication system for emergency and rescue operations should function independently of the availability of fixed infrastructure. However, the communication system should be able to exploit fixed infrastructure if such is available. The communication patterns will include both group and peer-to-peer communications.

Efficient and reliable communication is mission critical. The deployment time must be minimized. When the rescue team arrives on the scene of an incident, communication should be established and maintained with as little human interaction as possible.

The risk of intervention by malicious actors differs depending on whether the rescue operation takes place in a hostile or a benign environment. However, the communication system must function in either situation. Therefore, adequate protection must be built-in. Regulations and laws also impose security requirements.

The MANET may include nodes from a single security domain or multiple security domains. Operations cross national borders can be examples of the latter.

1.5.3 Network architecture

Different scenarios for how ad hoc technology will be used in future communication systems for emergency and rescue operations are discussed by Winjum in [105]. A probable application is as an extension to existing infrastructure-based systems, e.g., as shown in Figure 2.

The MANET may have a connection to fixed network through one or more gateways. The MANET internet connection may include a satellite link, a fixed or a mobile base station (possibly co-located with the on-site rescue management) or other. References [9] and [46] describe the connection of MANETs to the Internet with the
aid of Mobile IP assuming a reactive or a proactive ad hoc routing protocol, respectively.

Ad hoc networks could be used as transit networks. Throughout this thesis MANETs are assumed to be stub networks.

![Network Architecture](image)

**Figure 2 Network architecture**

Other sub groups of ad hoc networks may also have applicability for future disaster relief communication. Sensor nodes and sensor networks will likely come to play an important role. *Wireless sensor networks (WSNs)* are a type of ad hoc networks with different requirements and features than MANETs. Data are routed towards a base station. The nodes are normally static and more resource constrained than traditional MANET nodes. WSNs can be formed through mass deployment of a huge number of sensor nodes.

*Vehicular ad hoc networks (VANETs)* may also be exploited in disaster relief communication, but the role is less evident. In VANETs the nodes move along roads. The nodes communicate vehicle-to-vehicle or via base stations and fixed infrastructures available along the road. The node mobility is expected to be higher than in traditional MANETs, and the size of the VANET may be very large (potentially including all vehicles in the world).

Throughout this thesis ad hoc networks and ad hoc technology refer to mobile ad hoc networks unless other is specified. Other types of ad hoc networks are considered out of scope of this dissertation.
1.5.4 A note on MANETs and TETRA

Norway is currently (2007) planning to implement a TETRA (Terrestrial Trunked Radio) -based emergency network over the next years. TETRA is an open standard for emergency communication defined by the European Telecommunications Standards Institute (ETSI). The TETRA system is infrastructure-based, and has limited capacity. The emergency network for Norway is specified to cover only approximately 80% of the land area, and, hence, 90% of the population. MANETs could be combined with TETRA in order to provide higher data capacity, extend the coverage area and reduce the load on base stations.

TETRA uses Time Division Multiple Access (TDMA) technology, and offers 4 user channels (4 TDMA timeslots). The number of concurrent flows within one cell is hence low. The maximum data rate offered is 28.8 Kbps (7.2 Kbps per channel). This is in Trunked Mode operation (TMO) where the communication is set up with the aid of a TETRA base station. The TETRA terminals can also communicate directly or via another node in the Direct Mode operation (DMO). DMO can be used if the nodes are out of range of a TETRA base station or the base station for other reasons is not accessible. But TETRA DMO offers limited functionality regarding data rate, range (number of hops) and concurrent flows compared to what can be obtained with a mobile ad hoc network. The maximum data rate in DMO is 7.2 Kbps. TETRA 2 extends the data rates to some hundred Kbps. However, TETRA 2 requires a significantly higher density of base stations than TETRA in order to cover the same area. (Figures from Belgium indicate a factor of approximately 100 [105]). TETRA combined with ad hoc technology may be a more viable solution.

Our research has focused on mobile ad hoc networks and has sought to be neutral to specific implementations and possible co-existing technologies. The network architecture in Figure 2 has been assumed throughout the work.

1.5.5 Summary

Whereas some ad hoc networks may be truly ad hoc, mobile ad hoc networks for operational scenarios such as emergency and rescue operations will have a planned origin. Common points of trust will exist, and pre-configuration of security parameters will be possible.
Mobile ad hoc networks are expected to play an important role in future communication systems for emergency and rescue operations. However, exactly how the ad hoc technology will be used is yet to be concluded. The mobile ad hoc network should operate independently from fixed infrastructure, but be able to exploit such infrastructure if available.

The wireless media and nodes operating both as routers and communication end-points make the ad hoc network layer routing information prone to attacks. Hence, a protected routing protocol is a prerequisite for a reliable network service.

### 1.6 Threats to ad hoc networks

#### 1.6.1 Threats: MANETs versus infrastructure based wireless networks

Basically, MANETs are exposed to the same types of threats as other wireless mobile communication systems. Attacks can be directed towards any layer of the protocol stack. Jamming, eavesdropping, replays, falsification, insertion and modification of messages are inherently easy. Battery-powered terminals imply vulnerability to energy-depleting attacks.

However, MANET communication does not suffer from failures in fixed infrastructures such as access points and base stations, transmission network and core network in the same way as the infrastructure-based wireless and wired communication systems do. On the other hand, whereas the routers in infrastructure-based networks are part of an established and (at least to some extent) controllable infrastructure, the same is not true in ad hoc networks. The nodes acting both as routers and communication end-points make the ad hoc routing protocols more prone to attacks.

#### 1.6.2 Threats to MANET routing protocols

Attacks are often classified as either active or passive. A passive attack can be defined as an attack where an adversary attempts to defeat a cryptographic technique by simply recording data and thereafter analyzing it. An active attack involves an adversary who modifies or injects messages [63].
Passive attacks are mainly threats against communication privacy and anonymity, rather than against the network’s function or routing protocol [39]. Active attacks constitute the major threats. That is, passive eavesdropping on routing messages could be used as part of traffic analysis, which could be exploited to launch other attacks. But this is assumed not to represent any significant threat in the emergency and rescue operations scenarios.

Attacks on ad hoc routing protocols generally fall into one of the two categories; routing-disruption attacks and resource-consumption attacks [39]. Routing-disruption attacks refer to attacks where the attacker attempts to cause legitimate data packets to be routed in dysfunctional ways. In resource-consumption attacks, the attacker injects packets into the network for the purpose of consuming bandwidth or exhausting the nodes by depleting their energy resources or using up memory. In general, from an application-layer perspective, attacks on the ad hoc routing protocol appear as various types of Denial-of-Service (DoS) attacks. Nodes that perform active attacks with the aim of damaging other nodes by causing network outage are considered to be malicious.

Surveys of attacks on ad hoc routing protocols are found in [22], [23], [39], [40] and [94]. Related information is also found in [6], which discuss generic threats to routing protocols for wired networks. Examples of attacks on routing protocols include:

**Bogus routing information:** Successful insertion of bogus routing information may lead to corrupt routing tables.

**Black holes** refer to nodes that attract traffic by maliciously advertising shortest path to other nodes. The black holes may choose to forward routing information correctly but discard other data. **Grey holes** selectively forward data.

**Wormholes** [42] refer to adversary nodes colluding by tunneling packets from one part of the network to another, making geographically distant nodes falsely believe that they are neighbors. The wormhole nodes may then act as black holes or grey holes.

**Invisible node attacks** [61] are basically wormhole attacks consisting of only one node that passively relays routing messages. The invisible node can make 2-hop nodes falsely believe they are 1-hop neighbors. The attack can be set up with the aid of directional antennas.
**Replays** are retransmissions of earlier messages for the purpose of corrupting routing tables or exhausting the channel or other nodes.

**Rushing attacks** [41] are targeted at reactive routing protocols where routes are established on demand and in which duplicate messages are discarded. The attacker distributes route requests quickly through the network, *e.g.*, by extending the transmission range through increased antenna gain or augmenting the transmission power. Legitimate requests that arrive later are discarded.

**Byzantine behavior** refers to nodes that do not behave according to the protocol. This includes insertion of false or modified routing messages, masquerade, selfish behavior, delayed response and other deviations from the protocol. Malicious nodes may also try to cause routing loops.

**Selfish nodes** give priority to own traffic. They are reluctant to forward data for other nodes, and prefer to save their battery for own communications. (It could be argued that this represents a kind of passive attacks, but we regard it as an active attack where the node actively chooses not to cooperate in accordance with the protocol.)

**Impersonation** (*masquerade* or *spoofing*) can be launched by replays or old routing messages or insertion of false routing information for the purpose of corrupting routing tables, exhausting nodes or gain unauthorized access to the network.

A **Sybil attack** [25] refer to a single node capable of presenting multiple identities. This can be exploited in impersonation attacks.

**Blackmail:** In ad hoc network routing protocols where the nodes attempt to keep track of perceived malicious nodes in a “blacklist” at each node, *e.g.*, as in *watchdog* and *pathrater* [62], an attacker may blackmail a good node, causing other good nodes to add that node to their blacklists, thus avoiding that node in routes [40].

### 1.6.3 Threat model

The classical Dolev-Yao threat model [24] assumes that the adversary may eavesdrop on any message, modify and replay messages, and forge source and destination addresses. But the adversary cannot produce a valid signature on other nodes’ behalf or decrypt a message that has been encrypted with a key she does not possess.
For ad hoc networks the threat model must cope with one or a few compromised nodes. The nodes are out in the field. The physical protection of the communication nodes can be limited. The probability of one or more compromised routers (nodes) will thus expectedly be greater than in traditional wired and wireless networks. That is, the threats may come from *external attackers (outsiders)* as well as *internal attackers (insiders)*. However, the number of internal attackers will expectedly be low. Internal attackers here refer to nodes with access to secret key material. External attackers are adversaries without access to any special information other than what can be obtained by listening to the wireless transmissions. Attacks can be launched both by nodes that deliberately act maliciously, and nodes that unintentionally do not behave according to the protocol, *e.g.*, due to a hardware or software error.

A somewhat more complex attacker model for ad hoc networks is used in [39] and [40]. The model assumes an attacker consists of more nodes. Both external nodes and compromised nodes are included. The attacker owns all the cryptographic key information of compromised nodes and distributes it among all its nodes. The attacker can be passive or active, and is characterized by the number of nodes it owns in the network and the number of nodes it has compromised. An *Active-n-m* attacker is an active attacker consisting of *m* nodes that has compromised *n* good nodes.

### 1.6.4 Summary

Ad hoc routing protocols are more prone to attacks than their wired equivalents. Threats may originate from external attackers as well as insiders. An unprotected network service is vulnerable to active attacks. Active attackers may cause route-disruption and node exhaustion. Passive eavesdropping is considered not to be harmful.

Our work assumes that security schemes for ad hoc routing protocols should resist attacks from externals as well as a limited number of internal nodes.

### 1.7 A summary of the thesis

A reliable ad hoc network service requires protected routing control traffic. Authenticated routing messages, *i.e.*, integrity protected and source authenticated messages, are demanded. A proper key management system is an essential prerequisite of any successful cryptographic protection scheme. A natural starting
point was thus to survey existing key management schemes proposed for ad hoc networks (Paper I in Part II) [34]. The aim was to discover any dominating or more suitable schemes. The proposals were evaluated according to a set of evaluation criteria covering applicability under various network assumptions, security, robustness, scalability and simplicity.

The assumed threat model means that survivability despite one or more compromised nodes is required. Asymmetric schemes enable unique identification and authentication of the sources of route management traffic, and can also be used for verification of message authenticity. These are desirable features. If identifiers that are by default sent as part of the routing messages can be used as public keys, intuitively, identity-based schemes may scale better than traditional certificate based schemes. This observation led to a closer analysis of the implications of identity-based signature schemes for the protection of ad hoc routing information (Paper II in Part II) [33].

One of the conclusions was that signature sizes can be a more limiting factor for the applicability of asymmetric schemes than computational cost. Bandwidth efficient protection schemes are sought. That is, with resource constrained nodes, both little overhead and low computational cost is beneficial. This makes hybrid solutions interesting. A typical assumption for protection schemes is that all ad hoc routing messages must be signed. However, all messages may not need equally strong protection in order to provide a reliable network service. Especially in proactive protocols, a large number of the periodic routing messages primarily serve as “heart-beat” messages. They do not necessarily report any important changes. This was the background for the development of a hybrid protection scheme for the Optimized Link State Routing protocol OLSR [18] (Paper III in Part II) [32]. The hybrid protection scheme uses identity-based signatures and hash chains.

Practical communication systems will normally require a way to expel compromised and maliciously behaving nodes. This can be achieved through an appropriate revocation scheme. The revocation techniques proposed for fixed networks are generally not well suited for ad hoc networks. Some require guaranteed connectivity to a central trusted entity to check certificate status on-line. Others trade update cost for timeliness or vice versa. Schemes proposed for ad hoc networks typically rely on the ad hoc nodes’ exchanges of accusations or revocation lists obtained from trusted entities during periods with connection to the fixed net. Revocation due to false accusations may occur. The revocation lists typically cover all
revocations in the *entire* security domain, and not only the subset of revocations interesting for the nodes in a particular MANET.

Assuming that most practical ad hoc networks will have one or more gateways connecting them to external networks, the gateways could be exploited in the filtering and distribution of relevant revocation information. With a proactive routing protocol, the gateway will know which nodes are in the ad hoc network. This led to the development of the *MANET revocation list (MRL)* scheme (Paper IV in PART II) [31]. The MRLs are set up with the aid of information learned from the routing protocol and one or more trusted gateways with connection to a trusted revocation entity.

The work has concentrated on the network layer and the protection of routing information. A natural extension would be to investigate security on multiple layers and security as a cross-layer issue. As an example, black holes may be easier to detect by combining network layer routing information with connectivity experienced by the application layer. Common data-bases could be shared by multiple layers. Treating security as a multiple-layer and cross-layer issue, could lead to increased efficiency and possibly better, less fragmented solutions compared to what is obtained by adding security layer by layer.

We have assumed that compromised nodes should not be included in the ad hoc network, *i.e.*, a “Stop-at-the-network-layer” idea. However, some systems may require that the compromised nodes are included in the network in order to erase keys, *e.g.*, keys used to protect application data. An example of a scheme that utilizes key-zeroing messages is described by Jungels, Raya and Hubaux in [45]. It assumes that the keys are stored in tamper-proof devices. If the node has fallen into wrong hands, the key administrator deletes the key material through a deletion message acknowledged by the tamper proof device. This requires a network connection to the compromised node. On the other hand, such information could also be included in the transmissions in such a way the key material is deleted whenever the compromised node tries to eavesdrop on the communication.
2 Security in ad hoc networks

2.1 Ad hoc routing protocols

With the wireless media and mobile nodes acting as routers, ad hoc networks present different requirements than wired networks do. A large number of ad hoc routing protocols have been proposed. Some have received more attention than others, but there is no generally adopted standard. That is, there are ongoing standardization efforts. After classification of the routing protocols, the next sections summarize the standardization work, and outline some of the candidate standard ad hoc routing protocols.

Surveys of ad hoc routing protocols are found in [37], [56], and [59].

2.1.1 Classification of ad hoc routing protocols

A widely adopted classification of ad hoc routing protocols is according to their routing strategy, i.e., proactive (table-driven) or reactive (on-demand). In [37], Hong, Xu and Gerla extend this categorization to include the network structure underlying the routing protocols. Their categories include flat, hierarchical and geographic position assisted routing. Flat routing protocols are further divided into Reactive and Proactive.

Reactive routing protocols establish routes on demand. Proactive protocols maintain the routing tables irrespective of the actual use of the routes. Reactive routing protocols generally impose less overhead, but longer delays to establish a route than proactive protocols do. A drawback with proactive protocols is the constant bandwidth and energy consumption due to periodic updates. However, with scattered traffic patterns and high mobility, proactive protocols produce higher routing efficiency than reactive protocols do. The routes are known in advance. This is a desirable property also for real-time applications.

Hybrid routing protocols make use of both reactive and proactive approaches. As an example, in the Zone Routing Protocol (ZRP) [28] the nodes use a proactive approach within a specific radius and a reactive approach to reach nodes outside this zone.
When the network size increases beyond some threshold, the link and processing overhead of flat routing schemes become intractable. This motivates a hierarchical approach. Each node participating in flat routing protocols plays an equal role. In contrast, the nodes in hierarchical routing schemes are organized in groups, and nodes are assigned different functionalities inside and outside the group. The group leader node communicates with other nodes on behalf of the group. An example of hierarchical routing is the Internet hierarchy.

Flat routing protocols adopt a flat addressing scheme [37]. Some hierarchical protocols require a hierarchical addressing system. Hierarchical addressing schemes assume the nodes within an area to have related addresses, e.g., a common subnet address. In flat addressing schemes all bits in the addresses are used for each forwarding decision. The nodes need not have addresses from a specific range to be routable in the MANET.

Geographical routing assumes the nodes to be equipped with location-finding equipment such as the Global Positioning System (GPS).

Ad hoc routing protocols can also be classified according to other criteria such as routing metrics and whether the protocols use link-state or distance-vector routing algorithms. In distance-vector routing, the nodes maintain a view of the distance from themselves to all other destinations. This information is exchanged with neighbors. In link-state routing the nodes inform all other destinations in the network about their links to neighbors. A routing metric is any value used by routing algorithms to determine whether one route is superior to another [105].

2.1.2 Standardization work

The MANET working group (WG) of the Internet Engineering Task Force (IETF) has been established to standardize IP routing protocol functionality for ad hoc networks.

The Optimized link state routing protocol (OLSR) [18], the Ad hoc On-demand distance vector routing protocol (AODV) [86] and the Topology dissemination based on reverse-path forwarding protocol (TBRPF)[78] have earned status as experimental RFCs. Current Internet-drafts from the MANET WG by December 2006 comprised the Dynamic MANET On-demand (DYMO) routing protocol [13], OLSR version 2
Dynamic Source Routing Protocol (DSR) [44] and simplified multicast forwarding for MANETs (SMF) [60].

In addition, another IETF working group, the Open Shortest Path First (OSPF) WG, works with an adaptation of the OSPF routing protocol [19][69] for ad hoc networks, referred to as W-OSPF or OSPF-MANET. Two techniques dominate; the OSPF overlapping relays (OSPF-OR) [14] and OSPF-MANET designated router (OSPF-MDR) [77]. The OSPF WG has not yet (Jan 2007) reached consensus on which one of these approaches is the superior.

OLSR, OLSR version 2, TBRPF, and the OSPF-MANET protocols [14] [77] are all proactive link-state protocols. AODV, DSR and DYMO are reactive distance-vector protocols. The MANET WG currently (Jan 2007) pursues one proactive (OLSR version 2) and one reactive (DYMO) routing protocol. The official charter maintained by the IETF secretariat indicates that the MANET WG may consider a converged reactive and proactive approach.

2.1.3 Proactive ad hoc routing protocols

2.1.3.1 OLSR and OLSRv2

In the Optimized link state routing protocol (OLSR) [18] and OLSR version 2 (OLSRv2) [17], distribution of topology information is optimized compared to classical flooding through the use of Multipoint Relay (MPR) nodes. Only nodes that are chosen as MPRs forward routing information. Each node selects MPRs from its 1-hop neighbors in such a way that all 2-hop neighbors are covered by at least one MPR.

OLSRv1 specifies four message types: HELLO, Topology Control (TC), Multiple Interface Declaration (MID) and Host and Network Association (HNA). The HELLO messages are used for link sensing, neighbor detection and MPR selection, and are not forwarded by the MPRs. The other message types are flooded through the MPR nodes. TC messages convey topology information, and are only emitted by MPR nodes. MID messages are sent by nodes with more than one OLSR interfaces. HNA messages announce non-OLSR interfaces. The nodes are uniquely identified by their main address included as the originator address in every routing message. Routing tables are computed from the information exchanged through the HELLO and TC messages. 1-hop and 2-hop neighbors are detected from the HELLO messages.
Information concerning the 3-hop and more distant nodes are found from the TC messages.

OSLRv2 only specifies HELLO and TC messages, and adopts the generalized MANET packet and message formats described in [16]. Information such as gateway announcements is included as Type-Length-Value fields (TLVs) in TC messages. Thus, the generation of TC messages in OLSRv2 is not limited to MPR nodes.

Old or duplicate routing messages are detected with the aid of the message sequence numbers, and are discarded. Each node increments the message sequence number for every new message it generates.

2.1.3.2 OSPF-MANET protocols

OSPF-MANET provides reliable flooding. All routing messages are acknowledged. Backup nodes forward the routing control traffic if the expected forwarding node fails to do so.

The proposed OSPF-MANET solutions; OSPF overlapping relays (OSPF-OR) [14] and OSPF-MANET designated router (OSPF-MDR) [77], differ in the way they optimize the flooding of routing information. OSPF-OR nodes choose overlapping relays (ORs) parallel to the MPRs in OLSR. Non-OR nodes act as backup ORs that retransmit the routing messages if the ordinary OR fails to do so. OSPF-MDR use MANET designated routers (MDRs) to flood routing information. The decision to become a MDR or a backup designated router (BMDR) is made by the nodes themselves.

Topology information is disseminated through link-state update (LSU) packets containing link-state advertisements (LSAs). The other packet types; HELLO, database description (DD), link-state request, and link-state acknowledgement are not forwarded outside the 1-hop neighborhood. HELLO messages are used for neighbor detection and link sensing. Link-state acknowledgements are used to locally acknowledge the reception of LSAs. DD and link-state requests are used to synchronize the link-state databases between adjacent nodes.

The OSPF-OR and OSPF-MDR extensions to the OSPF packet format are carried in a link local signaling (LLS) data block attached to HELLO and DD packets. OSPF-OR in addition specifies a new LSA type – link LSA that is used to distribute information about 2-hop neighbors.
OSPF-OR and OSPF-MDR packets are encapsulated in IP. The nodes are uniquely identified by their router id included in each routing packet.

2.1.4 Reactive ad hoc routing protocols

2.1.4.1 DSR

In the Dynamic Source Routing protocol (DSR) [44] routes are established on demand. Route requests are flooded into the network. Intermediate nodes append their addresses to the route request messages as the queries travels through the network. The target of the route request returns a route reply to the initiator containing a copy of the accumulated route recorded in the route request message. This “source route” is then included in packets sent to the target. The source route defines which hops the packet is to traverse. Packets are acknowledged hop-by-hop. If a link breaks, a route error message is returned to the source nodes that use this link.

2.1.4.2 AODV and DYMO

The Dynamic MANET On-demand (DYMO) [13] routing protocol is a descendant of the Ad hoc on-demand distance vector routing protocol (AODV) [86] and DSR [44] designs. In AODV and DYMO routes are discovered on demand by flooding a route request (RREQ) into the network. During this dissemination process, each intermediate node records a route to the originating node. The destination (or intermediate nodes with a valid route to the destination) unicasts a route reply (RREP) along the path that the RREQ was received. Each node that receives the RREP records a route to the target node, and unicasts the RREP towards the originating node. In DYMO, only the destination node initiates RREPs.

Differently from AODV, DYMO also allows intermediate nodes append additional routing information to the routing messages. That is, if the nodes believe that this information will alleviate future RREQs. Other nodes may remove the additional routing information if they consider it outdated.

Each destination in the routing table is recorded with its latest known destination sequence number (DYMO calls it DYMO sequence numbers). All nodes increment their sequence numbers before issuing a new RREQ. The destination sequence numbers ensure no routing-loops. The routing tables are not updated unless a route with a higher sequence number is detected. That is, of two routes with same
destination sequence number, the one with the lowest HopCount is chosen. The HopCount is incremented by each node that the routing message traverses, and indicates the number of hops from the originator of the message.

The nodes maintain their routes and monitor links over which traffic is moving. A route error (RERR) is sent to the packet source if the current route is broken.

AODV nodes should use an expanding ring search technique to prevent unnecessary network-wide dissemination of RREQs. This means that the flooding scope of RREQs is gradually increased until a RREP is received. The range of dissemination is controlled by the TTL (Time To Live) field of IP header. Expanding ring search is optional in DYMO.

AODV and DYMO messages are encapsulated in UDP over IP. The nodes are identified by their IP addresses.

2.2 Protecting the ad hoc routing protocols

2.2.1 Trust model

A typical assumption for the protection of ad hoc routing information is that nodes in possession of a valid secret key are trustworthy. Others are not. The key can be a group key established through the collaborative effort of all or a subset of the MANET nodes, or a key distributed from a trusted third party (TTP) or approved by a TTP, e.g., through a certificate signed by the TTP.

RFC 3756 [71] discusses trust models for various networks, including ad hoc networks. The RFC assumes truly ad hoc networks where the nodes meet for the first time and no prior trust relations exist. It assumes that there is no guarantee that other nodes behave according to the protocol. But with IP addresses derived from public keys, the nodes can trust that they communicate with the same node all the time. That is, the node picks a private key and calculates the corresponding public key. The IP address is then calculated from the public key. Others cannot easily steal the IP address. Only the owner of the private key will be able to produce a signature that can be verified with the public key, and hence, prove the ownership of the address.

The trust model for disaster relief MANETs assumed in our work, is related to the hierarchical organizational structure depicted in Figure 1. A Key Administration
Centre (KAC) or trusted entity is assumed in each security domain. The KAC may be a national key administration centre related to a national rescue co-ordination centre.

All the nodes within the security domain trust the KAC to conceal its secret keys and to protect them from unauthorized access. In case of a traditional public key scheme, the KAC is believed to issue certificates only to authorized nodes. The KAC is also trusted to generate secret keys for the other nodes in the security domain, and to distribute these in a secure way (through a confidential and authenticated channel).

Ordinary nodes in possession of a secret valid key are trusted to operate according to the agreed protocol and forward data correctly. Nodes from other security domains may be trusted in the same way, e.g., as a result of the KACs of the involved security domains have cross-certified each other. The ordinary nodes are in addition trusted not to disclose their secret key to unauthorized parties. Hence, the assumed trust hierarchy includes ordinary nodes and the KAC. For practical reasons, key material can also be distributed through nodes that are more protected and trusted than the ordinary ones, e.g., nodes at the local co-ordination centre. This means an additional level between ordinary nodes and the KAC in the hierarchy of trust. These “distribution nodes” can be trusted to store key material for other nodes in a secure way, and only let authorized nodes receive keys –after proper identification. Both the KAC and “distribution nodes” are typically located outside the MANET.

2.2.2 Definition of security for ad hoc routing protocols

The general terms used to describe security; availability, integrity, authentication, confidentiality and non-repudiation apply to MANETs as to other networks. However, every factor is not equally important for the protection of ad hoc routing information. Nodes can be trusted to forward data correctly without necessarily being authorized to gain access to the information contained in the transferred data. Ordinary payload is typically end-to-end and can be protected with any end-to-end security system like IPsec [50]. Routing messages are sent to immediate neighbors, processed, possibly modified, and re-transmitted [110].

Availability is a number one concern. The transmission of ordinary payload can assume an already running network service. The routing messages are a necessary precondition for the establishment and maintenance of routes that makes communication
of application data possible. Integrity and authenticity is important both for protection of payload and for the establishment of reliable routes through authorized nodes. However, whereas confidentiality of the application data usually is a primary concern, hiding the contents of routing messages is normally not necessary. Furthermore, whereas several applications may require non-repudiation for log and audit, routing information has basically only instant value.

The routing messages demand end-to-end protection. Reactive protocols such as AODV and DYMO require special attention to the HopCount field that is modified by intermediate nodes. In contrast, the proactive OLSR and OLSRv2 make no use of mutable fields such as Hop Count and TTL in the route calculation. These fields need not be protected. On the other hand, neighbor detection and topology changes are of great concern.

With malicious nodes in the network, route replies from intermediate nodes instead of the destination node in AODV are a security risk. The intermediate nodes’ inclusion and removal of additional routing information in the DYMO routing messages may also represent a security challenge.

2.2.3 Basic protection mechanisms

Whereas the security schemes proposed for ad hoc routing protocols are different, the designs generally rest on one or more methods from the same set of primitives. The next sections describe the most common basic protection mechanisms. These mechanisms are prerequisites for the schemes in sections 2.2.4 and 2.2.5.

2.2.3.1 Message authentication codes

Message authentication codes (MACs) are a candidate type of mechanism for integrity protection and authentication of routing message origin.

A MAC algorithm is a family of functions that take a secret (symmetric) key and an input text of arbitrary length, and produce a fixed length MAC as output [63]. Given the key and the text, the MAC is easy to compute. Without knowledge of the key, it is computationally infeasible to compute a valid text-MAC pair.

On reception of a routing message with a corresponding MAC, nodes in possession of the correct key can calculate the MAC of the data and verify that it equals the received MAC. In this way, integrity and authenticity can be checked.
The MAC length is typically short – 32 or 64 bits are common choices. The small sizes plus low computational cost are important advantages.

MACs can be generated in different ways. Both stream-ciphers and block-ciphers may be used. One possibility is to use a block-cipher in cipher-block-chaining (CBC) mode. The Advanced Encryption Standard, AES [72] has replaced the Data Encryption Standard (DES) [73] as a generally adopted standard that can be used to generate CBC-MACs. Other modes of operation may also be used. Reference [27] describes Galois Message Authentication Codes (GMAC) generated with the aid of AES in Galois/Counter mode of operation for IPsec’s [50] ESP [49] and AH [48]. Message authentication codes can in addition be generated with the aid of keyed hash functions. An example is HMAC [54].

2.2.3.2 Cryptographic signatures

Stinson [99] defines a signature scheme as a method of signing a message stored in electronic form. The signature scheme consists of two components; a signing algorithm and a verification algorithm. If a message is signed with the private signing algorithm, the resulting signature can subsequently be verified using the public verification algorithm. Cryptographic signatures can be used for data integrity, authentication and non-repudiation.

Cryptographic signature schemes can be classified as signature schemes with appendix or signatures schemes with message recovery. Signature schemes with appendix require the original message as input to the verification algorithm [63]. Signature schemes with message recovery do not. The original message is recovered from the signature itself. The latter may give shorter messages, but the contents cannot be evaluated before the message has been recovered.

Protection schemes for ad hoc routing information normally rely on signature schemes with appendix. It is beneficial to be able to evaluate the routing message contents for further processing without first having to spend energy on the recovery. Cryptographic signatures with appendix are assumed in the further.

Cryptographic signature schemes include both traditional certificate-based digital signatures and identity-based signature (IBS) schemes [95]. IBS schemes use the identifiers, e.g., IP addresses, as public keys. The private keys are provided by a
trusted third party, and are derived from the identifier (public key) of the node plus the private master key of the trusted entity.

RSA [92] and DSA [74] are representatives of well-known digital signature schemes. Shamir’s original IBS scheme [95] was also derived from RSA. Newer IBS schemes include Cha and Cheon [12], Hess [35], Paterson[83], and SOK[93] rely on the Weil and Tate pairings on elliptic curves.

Signature schemes are almost always used in conjunction with a very fast public cryptographic hash function. MD5[91] and SHA-1[76] have been much used hash functions. The collision-free hash function takes a message of arbitrary length and produces a message digest of a specified size (160 bit is a popular choice) [99]. The message digest is then signed using the signature algorithm. Relying on discrete log and factoring problems, signature algorithms are known to be computationally expensive. The signatures are also generally large; 1024 and 2048 bits are common RSA signature sizes. DSA signatures are shorter. Typical sizes are 160 – 224 bits.

2.2.3.3 Hash Chains

Hash chains [55] have been proposed for the protection against short distance frauds and sequence number deceptions in a number of secure routing protocols. It has also been proposed to prevent topology information from being distributed further than a specified number of hops (SLSP [80]).

Hash chains are created with the aid of a cryptographic hash function. If a hash function is to be considered secure, it requires preimage resistance (one-way property), second preimage resistance (for a given input it is hard to find another that gives the same hash value) and collision resistance (hard to find pairs of different inputs that return the same hash value). However, hash chains only depend on the one-way property.

Hash chains are constructed by repeated hashes of an initial random seed $RND$: $h_1=h(RND)$, $h_2=h(h_1),...,h_{n-1}=h(h_{n-2})$, $h_n=h(h_{n-1})$ ($h()$ is the cryptographic one-way hash function). If the last value in the hash chain, i.e., the hash anchor $h_n$, is distributed through an authenticated channel, the receiver can verify whether later disclosed values originate from the same hash chain or not. Repeated hashes of the received value should return the hash anchor. Forward hash calculation is fast and easy. The one-way property makes it hard for anyone but the creator to find the preceding values in the hash chain.
After SHA-1 was reported broken in 2005 [103], there is a move towards other members in the SHA family such as SHA-256, SHA-384 and SHA-512. However, the reported attack concerned the collision resistance and not the one-way property. SHA-1 may still be applied for hash chains.

2.2.3.4 Redundancy

Redundancy increases the robustness against malfunctioning and selfish nodes and nodes that for other reasons exhibit Byzantine behavior. Examples include the OSPF MANET protocols [14] [77] that have backup nodes that forward the routing messages if the ordinary overlapping relays or MDRs fails to do so. The self-healing community approach in [52] extends the responsibility of backup forwarding to all 1-hop neighbors that the strict 2-hop neighbors have in common.

Redundancy intrinsic in the routing protocols in the form of periodic refresh and route maintenance message also increase the resilience. As an example, even with signed routing messages and adversaries capable of sporadically forging a message signature, the protocol will only be disturbed temporarily. Acknowledged forwarding of routing messages as in the OSPF MANET protocols and DSR represents “redundancy” that increase the reliability of the routing protocol.

2.2.3.5 Reputation schemes

Reputation schemes have been proposed to thwart selfish nodes and other types of Byzantine behavior. Reputation schemes typically rely on monitoring of node behavior and accusations. The nodes listen to the transmissions of routing packets in promiscuous mode. Bad behavior can be detected by keeping a copy of the packets and comparing them to the retransmitted versions. If node hears another node retransmit the packet with modified contents, it can warn the others. Examples of reputation-based schemes include CONFIDANT [10] and Watchdog and Pathrater [62] (sections 2.2.4.2 and 2.2.4.3).

Assuming signatures are used to distinguish legitimate members of the network from unauthorized ones, reputation schemes can be used as an additional measure to counter misbehavior by insiders.
2.2.3.6 Message identifiers

The ability to uniquely identify routing messages is important for protection against routing loops and replay attacks. The node addresses are normally part of the identifier. In addition, most commonly, sequence numbers are used. The secure routing protocol ARAN [94] uses nonces and timestamps in the message identification. A nonce is a value used no more than once for the same purpose [63]. Sequence numbers represent one type of nonces.

Sequence numbers included in the routing messages enables the recipients to evaluate the freshness of the message. Old messages and duplicates of new messages that have already been processed are discarded. This prevents routing loops. It also protects against replay attacks. However, the sequence numbers need to be protected by a signatures or a MAC. Otherwise, the built-in suppression of duplicates and old messages could easily be exploited in DoS attacks. It is also necessary to keep a record of the latest sequence numbers received.

In [107], Winjum et al. propose acknowledged sequence numbers combined with message authentication codes are for the protection against replay attacks in OLSR. Adjih et al. [2] and Hafslund et al. [29] use sequence numbers, signatures and timestamps for the same purpose.

2.2.3.7 Packet leashes

Hu, Perrig and Johnson [42] propose packet leashes to counter wormhole attacks. A leash is information that is added to a packet to restrict the packet’s maximum allowed transmission range. Strict timing or localization information is a prerequisite. A geographical leash includes information about the locations of the nodes. Temporal leashes include a timestamp that enables the receiver to evaluate whether the packet has traveled longer than the maximum transmission distance (assuming the speed of light). Temporal leashes require stricter time synchronization than geographical leashes. It may be hard to distinguish a delay due to an additional travel distance from that introduced by lower layers’ normal media contention. Signing the temporal leash at transmission time makes it even more difficult. The time required for the signing process normally exceeds the transmission time with several orders of magnitude. Geographical leashes combined with a signature scheme ease detection of adversaries claiming to reside at multiple locations simultaneously.


2.2.3.8 Use of the protection mechanisms

Cryptographic signatures and message authentication codes are used to prevent successful insertions of false or modified routing information, black and grey holes, masquerade/spoofing and externals exhibiting Byzantine behavior. MAC and signatures can also be combined with sequence numbers (or time stamps) to protect against replay attacks and routing loops. Wormholes necessitate additional measures, e.g., packet leashes [42]. Redundancy is important for robustness and survivability. Hash chains can amongst other be used in the protection of mutable fields of routing messages in reactive protocols.

MACs and signature schemes are the fundamental mechanisms used in most protection schemes for routing information. The generic features of symmetric and asymmetric methods are therefore discussed in more detail in the sequel.

Table 1 provides a simplified overview and comparison of some of the attributes of symmetric and asymmetric methods that are relevant for protection of ad hoc routing information. The “+” sign means the scheme meets the requirements or has an advantage regarding this attribute. Drawbacks and missing features are marked “-”. The “+/−” indicate the method can meet the requirements, but not necessarily very easily.

<table>
<thead>
<tr>
<th></th>
<th>Symmetric methods</th>
<th>Asymmetric methods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAC</td>
<td>Digital Signatures</td>
</tr>
<tr>
<td>Computational complexity</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Overhead</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>No certificates</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Intrusion tolerance</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Source authentication</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Revocation</td>
<td>−</td>
<td>+/−</td>
</tr>
</tbody>
</table>

Computational complexity and overhead: Symmetric schemes are superior to the asymmetric ones in computational efficiency. For the same level of security they also
provide less overhead. The computational cost and overhead of asymmetric schemes are a challenge. And, differently from symmetric schemes, the processing cost and bandwidth consumption also grow rapidly with increasing security levels. Bandwidth consuming certificate exchanges imposed by traditional asymmetric schemes comes in addition.

If identifiers that are by default sent as part of the routing messages can be used as public keys, identity-based public key signature schemes may scale better than the traditional certificate based schemes. That is, Shamir’s original identity-based signature scheme [95] expands the signature size with a factor of two compared to RSA for the same modulus. Newer IBS schemes such as Cha and Cheon [12], Hess [35], Paterson [83], and SOK [93] that are based on the Weil or Tate pairings, provide shorter signatures. Implementations with supersingular elliptic curves with an embedding degree of 2 give signature sizes comparable to RSA. Various choices of curves and embedding degrees impose different message expansions and computational cost. The signature sizes can to some extent be traded for computational complexity and vice versa. Further details are found in [51].

Source authentication and intrusion tolerance: Symmetric MACs generated with a group key can be used to distinguish authorized members of the network from unauthorized ones. But a group key precludes unique identification of the source node. All nodes in possession of the secret key can generate a valid MAC. A node can insert and modify messages on other nodes’ behalf. This makes the system more vulnerable to attacks from malicious insiders. Another disadvantage is that the system security is compromised by a single captured node.

Pairwise unique keys would provide better intrusion tolerance, but would also necessitate a separate message authentication code for each possible recipient. For routing messages that are flooded in the network, this scales badly. Besides, all possible recipients of the routing message may not even be known in advance.

Signatures enable unique identification of the message originators. The asymmetric nature prevents that nodes can undetectably modify or insert messages on other nodes’ behalf. This means a stronger protection against inside attacks compared to symmetric schemes relying on a group key. Asymmetric schemes also fulfil the requirement that one or a few compromised nodes should not cause a total break of the system security.
Revocation and exclusion of nodes: With systems relying on a symmetric group key, exclusion of a single node implies re-keying. A typical solution would include a key distribution centre that sends a new group key encrypted with unique keys to each of the remaining nodes. This is bandwidth consuming, and there is no guarantee that all nodes receive the new key in a timely manner. This represents a threat to network availability. Although still challenging, revocation is easier with asymmetric schemes. Nodes can be excluded without need for the remaining nodes to change their keys. However, revocation is known to be difficult in wired networks, and is even harder in the ad hoc environment. There are problems related to factors such as timeliness, overhead, and no guaranteed connectivity. Revocation and exclusion of nodes are discussed in more detail in section 2.3.3.

Conclusion: None of the methods in Table 1 is fully compliant with the requirements. The protection of ad hoc routing messages would benefit from an asymmetric component, but the complexity and amount of overhead introduced by such schemes need close attention. A proper revocation scheme also needs further consideration.

2.2.4 Basic security modules

Basic security modules here refer to frameworks or security protocols that can be applied in various secure ad hoc routing protocol designs. The modules typically include one or more of the basic security mechanisms. The distinction made between basic security modules and secure routing protocols is not very strict, and there is to some extent overlap between the two categories.

2.2.4.1 TESLA

TESLA [87] (Timed Efficient Stream Loss-tolerant Authentication) is an authentication scheme designed for source authentication in multicast streams in lossy channels. It relies on symmetric cryptography and delayed disclosure of keys. The keys are subsequent values from a hash chain. The key used at time \( i \) equals a hash of the key used at time \( i+1 \). The hash anchor must be sent in the initial packet, and is authenticated with the aid of a digital signature. The hash anchor enables the nodes to verify that later disclosed keys originate from the claimed source. To send an
authenticated packet, the sender computes a message authentication code (MAC) with a key that is secret at that point in time. The receiver stores the message until the key later is disclosed. The nodes must be loosely time synchronized and know the key disclosure schedule. Otherwise adversaries could forge messages with the aid of already disclosed keys.

The TIK (TESLA with instant key disclosure) [42] extension to TESLA enables key disclosure without delay. It requires accurate time synchronization between all communicating parties, and makes use of a Merkle hash tree construction [64] to commit to the keys.

2.2.4.2 CONFIDANT

CONFIDANT [10] (Co-operation of nodes: Fairness in dynamic ad-hoc networks) is a reputation-based scheme that aims at detecting and isolating misbehaving nodes. It works as an extension to a reactive source routing protocol, e.g., DSR. The nodes monitor the behavior of their one-hop neighbors. By keeping a copy of a packet while listening to the re-transmission of the next node, malicious actions such as changes in message contents and dropped packets can be detected. When a threshold number of suspicious events have been detected from a node, the monitoring node alerts its friend nodes about the suspicious behavior. (Detailed specifications for how to make friends are not provided.) If there is sufficient evidence that the node is malicious, it is omitted in the route discovery. The signatures on the alert messages prove the accusations’ trustworthiness.

2.2.4.3 Watchdog and Pathrater

Watchdog and Pathrater [62] are two extensions proposed for DSR. The method can also be used for other reactive source routing protocols. The approach is reputation-based. The watchdog identifies misbehaving nodes and the pathrater helps the routing protocol avoid these routes. The watchdog listens in promiscuous mode to check that the next node in the path forwards data as agreed. It requires that the watchdog module keep a copy of the recent packets, and compare each overheard packets with the stored versions. The source of the route is notified if misbehavior is detected. The pathrater module uses the information from the watchdog combined
with link reliability data to pick the most reliable route. The path should be free from misbehaving nodes.

A main problem with the watchdog and pathrater approach is the vulnerability to blackmail attacks.

2.2.4.4 TIARA

TIARA (Techniques for Intrusion Resistant Ad hoc Routing Algorithms) [90] is a set of design techniques to protect ad hoc networks against DoS attacks. The techniques include flow-based route access control (FRAC), multi-path routing, source-initiated flow routing, flow monitoring, fast authentication, and the use of sequence numbers and referral-based resource allocation. The FRAC mechanisms drop packets not destined to and originating from authorized destinations and sources of packet flows. Multi-path routing refers to the maintenance of redundant routes. When multiple paths exist between the source and the destination, source-initiated flow routing enables the source to specify which path to use. The source node periodically sends encrypted and digitally signed flow status message to the destination. This flow monitoring helps detect path failures. Fast authentication refers to the inclusion of a path label at a node specific secret location within the packet. The sequence numbers are proposed to counter replay attacks. Resource depletion is thwarted by limiting the maximum amount of network resources a node is willing to allocate to a specific flow of packets. Additional resources are only granted if “referrals” from a number of other trusted nodes can be presented.

The techniques provided by TIARA are protocol independent, but they require extensive changes to existing protocols in order to be successfully incorporated [4].

2.2.4.5 SAR

Security-aware Ad hoc Routing (SAR) [109] is a framework for the inclusion of security metrics into the route discovery process of reactive ad hoc routing protocols. The aim is to find a route with a quantifiable guarantee of security. The nodes can be assigned different trust levels. Only nodes with the correct level of trust are able to process the route requests and replies. This can be implemented by encrypting the routing messages with different group keys for each trust level.
The authors of SAR call the implementation of SAR with AODV $SAODV$ - Security-aware $AODV$. This should not be mixed with the Secure Ad hoc On-Demand Distance Vector $(SAODV)$ protocol proposed in [110].

2.2.4.6 TMR

Trust Metric Routing (TMR) [105] is a framework for protection of ad hoc routing protocols that operates in the setting of proactive link-state routing. It has some similarity with SAR [109]. It relies on pre-shared group keys. However, whereas SAR encrypts the routing messages and builds routes through nodes with equal levels of trust only, TMR allows routing cooperation between different security domains. The users may choose between trustworthy routes and ordinary routes.

2.2.5 Secure ad hoc routing protocols

The next sections present a representative subset of secure ad hoc routing protocols. Secure ad hoc routing protocols are here defined as protocols where the protection scheme comes as an integral part of the protocol. That is, security enforcement is not left to other layers of the protocol stack. Surveys of secure ad hoc routing protocols are found in [4] and [39].

2.2.5.1 Secure proactive ad hoc routing protocols

2.2.5.1.1 SEAD

The $secure$ efficient ad hoc distance vector routing protocol $(SEAD)$ [38] is based on the DSDV [85] routing protocol, which is a proactive distance-vector protocol. Each node periodically broadcasts information about its shortest routes to all other destinations. Updates may also be sent when a change in the metric for one or more destinations have been detected. The neighbors use this information to update their own routing tables. In SEAD a hash chain is used to ensure other nodes cannot retransmit the routing message with a higher sequence number than just received or with a shorter distance than currently received.

Authenticated hash anchors for all nodes in the network are a prerequisite. The hash chain is divided into a number of sub-chains. The length of each sub-chain
equals the maximum network diameter. A node advertises a route with a distance and a sequence number together with a value from a hash chain. The sequence numbers are used for indexing the sub-chains. (The lowest sequence number points to the sub-chain that has the hash anchor as its last value). HopCounts are used for indexing within the sub-chains. (HopCount zero points to the first value in the sub-chain). Hashing the received value a number of times derived from the sequence number and the advertised distance should return the hash anchor. This ensures other nodes cannot retransmit the routing message with a higher sequence number than just received or with a shorter distance than currently received.

In addition to protection of the HopCount, it is also necessary to authenticate the routing message origin. SEAD suggests a broadcast authentication mechanism such as TESLA [87] is used for this purpose. The problem is that TESLA demands synchronized clocks. Pairwise shared symmetric keys is proposed as an alternative that does not require synchronized clocks. But this benefit comes at the price of increased key-administration cost.

2.2.5.1.2 SLSP

The proactive secure link state protocol (SLSP) [80] secures the discovery and distribution of link-state information with the aid of digital signatures. It uses hash chains to prevent topology information from being distributed further than a specified number of hops. SLSP can be used as a stand-alone proactive protocol or combined with a reactive protocol into a hybrid routing framework. Certificates are distributed in public key distribution packets, or are attached to link state update packets. Certificate issuance is not part of the protocol. Though, [80] indicates that this service can be provided by a coalition of nodes with the aid of a threshold scheme, the use of local repositories of certificates provided by the network nodes or a distributed instantiation of a certificate authority (CA).

2.2.5.1.3 Security extensions to OLSR

A number of security extensions have been proposed for OLSR. Adjih et al. [2], Hafslund et al. [29] and Winjum et al. [107] address replay attacks. The Winjum et al. proposes sequence number receipts and routing messages protected end-to-end with
the aid of symmetric message authentication codes. The other two rely on timestamps. In [29], Hafslund et al. suggest the messages are authenticated hop by hop with the aid of digital signatures. Adjih et al. [2] propose end-to-end protection enforced through asymmetric signatures or symmetric message authentication codes. Raffo et al. [88] extends the scheme in [2] with geographical leashes in order to protect against wormhole attacks. SOLSR [36] also focuses on wormhole attacks, suggesting temporal leashes combined with digital signatures. In addition, SOLSR use hash chains for protection of the mutable TTL and Hop Count fields. However, these fields are not used in the OLSR route calculation, and there is basically no need to protect them.

Fourati, Agha and Claveirole [26] suggest OLSR routes secured with the aid of threshold cryptography. Each MPR chooses a private/public key pair. The MPRs flood their public keys into the network. In addition, each MPR unicasts a share of its private key to each of its 1-hop neighbors, encrypted with the public key of the neighbor node in question. The neighbors sign the TC messages from this node with the received share. When a threshold number of different signed versions of the same TC message have been received, the signature can be verified.

It is not evident how a node can authenticate the neighbor nodes and their public keys before handing out shares of its private key. Trusted third parties are omitted. System security is jeopardized by Sybil attacks [25].

2.2.5.2  Secure reactive ad hoc routing protocols

2.2.5.2.1  SAODV

The Secure Ad-hoc On-demand Distance vector (SAODV) routing [110] is a security extension to AODV. Integrity and authenticity is protected through signatures covering all non-mutable fields of the routing messages. SAODV also protects against short distance frauds. The hop count field that is modified by the intermediate nodes is authenticated with the aid of hash chains.

A hash anchor is included in the signed part of the routing message. The hash anchor is set to a random seed hashed a number of times equal to the maximum number of hops allowed. The current hash value (initially set to the seed) is included in hash field in the unsigned part of the message. Every time a node receives a RREQ
or a RREP it verifies the hop count by hashing the value found in the hash field a number of times equal to the max hop count minus the current hop count. The message is discarded unless the result equals the hash anchor. The hop count is incremented and the hash value is replaced with a hash of the value received before the message is retransmitted. Thus, an intermediate node cannot undetectably forward the routing message with a lower hop count than received from its predecessor. (A different SAODV is presented in [109]. SAODV there stands for Security-aware AODV and refers to the implementation of SAR [109] for AODV).

2.2.5.2.2 Ariadne

Ariadne [40] is derived from DSR. As in DSR, the intermediate nodes append their ID to the RREQ message before relaying it. In addition, both the source node and each node traversed add a message authentication code to the RREQs. The MACs are calculated with the aid of pairwise keys shared between the node and the target. Pre-shared or TESLA keys can be utilized. Alternatively digital signatures can be used. The approach enables the target to authenticate both the source of the request as well as all intermediate nodes. A per-hop-hashing technique is used to ensure that no intermediate nodes are left out; each node replaces a hash value field in the routing message with a hash of its ID and the hash value received. The result is included in the input to the MAC calculation.

The RREP includes an accumulated list of addresses from the request, and is protected through a MAC added by the message originator. Multiple RREPs may lead to redundant routes. Ariadne suggests monitoring of packet delivery, multiple routes and packets sent along different routes in order to detect and mitigate routing misbehavior. Routes are chosen on the basis of prior performance in packet delivery. Differently from the Watchdog and Pathrater, feedback on which packet were successfully delivered is obtained from the target. This renders the system less vulnerable to blackmail attacks from intermediate nodes. The feedback is obtained through an extra end-to-end network layer message or by exploiting cross-layer information.
2.2.5.2.3 SRP

The Secure Routing protocol (SRP) [79] is a lightweight security extension for routing protocols that broadcast route requests, e.g., DSR and the reactive routing component of ZRP. The solution is similar, but simpler and less secure than the approach of Ariadne [40]. There is no authentication of the intermediate nodes. The integrity and authenticity of the route requests and route replies are ensured with the aid of a single message authentication code added by the message originators. Pairwise keys shared between the source and target nodes are a prerequisite. Differently from Ariadne, SRP does not prevent unauthorized nodes to be included in the route.

In [61], Marshall, Thakur and Yasinsac points out that SRP is vulnerable to nodes that passively relay routing messages without adding their address in accordance with the protocol, i.e., invisible nodes or “wormholes” consisting of one node. Ramachandran and Yasinsac show in [89] that this problem is not limited to SRP, but a generic problem to any “out-and-back” route discovery protocol in which intermediate nodes fix the reverse path during the forward request.

2.2.5.2.4 ARAN

ARAN (Authenticated Routing for Ad hoc Networks) [94] assumes route discovery packets are protected end-to-end and hop-by-hop by digital signatures. The scheme is designed with AODV and DSR in mind. The originator of the message includes its certificate and signs the message. The neighbors verify the signature, add their own certificate and sign the message before it is retransmitted. The next nodes that receive the message verify the signature with the given certificate. The signature and certificate appended by the successor is replaced with this node’s certificate and signature before the message is retransmitted.

Differently from SRP[79] where the MAC can only be verified by the end-nodes that possess the pairwise secret key, the digital signatures enable the intermediate nodes to verify the message origin. Furthermore, the end-to-end plus hop-by-hop digital signatures ensure routes are established through authorized nodes only. That is, as the intermediate signatures are changed hop-by-hop and not just appended as in Ariadne [40], there is no end-to-end authentication of the intermediate nodes. The nodes must trust the others to authenticate their neighbors correctly. On the other
hand, the hop-by-hop signatures represent a lower bandwidth cost than if all signatures were appended.

2.2.5.2.5 SPAAR

The Secure Position Aided Ad hoc Routing (SPAAR) [11] protocol take a different approach than the other protocols surveyed here in that location information is used to make routing decisions. A necessary precondition for SPAAR is that each device is able to determine its location. The routing tables contain information about node identities, location, transmission range and velocity. The routing messages are encrypted. The position information is thus hidden from unauthorized nodes. Time stamped sequence numbers prevent successful replay attacks. In order to counter wormhole and invisible node attacks (described in section 1.6.2), routing messages are only accepted from verified 1-hop neighbors.

SPAAR relies on asymmetric cryptography. The neighbors are authenticated with the aid of their public (authentication) key and certificate signed by a common trusted third party in advance. Each node in addition generates a private/public group encryption/decryption key pair for concealing the contents of RREQ and other forward messages. The public group decryption key is distributed to all verified neighbors encrypted under their respective public (authentication) keys. RREQs are encrypted hop by hop with the aid of the group encryption key of the current forwarding node. The authentication keys are used on the reverse path; RREPs are signed by the originator and encrypted hop-by-hop with the public (authentication) key of the next forwarding node.

The hop-by-hop protection has resemblance with ARAN [94]. Whereas ARAN only provides integrity and authenticity, the SPAAR approach provides hop-by-hop confidentiality in addition. However, as specified in [11], the RREQ in SPAAR is encrypted but not signed by the originator. Differently from ARAN, the authenticity of origin of the RREQs can thus not be verified end-to-end. It is different with the reverse RREPs. These are both signed by the originator and encrypted hop-by-hop.
2.2.6 Security extensions proposed in the MANET WG and OSPF WG Internet-Drafts and RFCs

The secure ad hoc routing protocols presented in section 2.2.5 all assume that the cryptographic protection scheme is an integral part of the routing protocol, but the Internet-Drafts and RFCs from the MANET WG and the OSPF WG described in section 2, generally suggest security is added by the lower layers.

OSPFv3 [19] relies on IPsec [50]. OSPF-MDR [77] provides no additional security specifications. OSPF-OR [14] suggests the ability to connect to the MANET is controlled by layer 2 security mechanisms such as IEEE 802.11i [43]. The AODV RFC [86] and DYMO Internet-draft [13] propose IPsec authentication header (AH) [48] as an appropriate authentication mechanism if the nodes share an appropriate security association. IPsec AH is also proposed in the TBRPF RFC [78] to counter impersonation attacks, and to prevent unauthorized nodes from joining the network via neighbor discovery. In addition, IPsec Encapsulating Security Payload (ESP) [49] is suggested for confidentiality.

The OLSR RFC [18] and OLSRv2 Internet-draft [17] suggest that digital signatures and other required security information are transmitted as separate OLSR messages in order to allow “secured” and “unsecured” nodes to co-exist in the same network. IPsec AH is also proposed to establish the authenticity of the OLSR control messages. However, it is pointed out that all OLSR control messages are point-to-multipoint transmissions and all receivers must be able to validate the authenticity. The OLSR RFC and OLSRv2 Internet-Draft suggest OLSR control traffic encrypted with the aid of Pretty Good Privacy (PGP) [5] or a shared secret cryptographic technique in cases where the network topology needs to be concealed.

The SMF Internet-draft [60] says authentication mechanisms should be considered to identify the source of an option header to reduce vulnerability to a variety of attacks. Further specifications are not provided.

2.2.7 On the applicability of IPsec and IEEE 802.11i for ad hoc networks

Routing messages are typically encapsulated as UDP over IP, e.g., OLSR, or sent as IP datagrams, which is the case with the OSPF-MANET protocols. Protection could be an integral part of the routing protocol, or enforced through lower layer protection schemes such as IPsec [50] or IEEE 802.11i [43]. This is illustrated in Figure 3.
The figure indicates end-to-end protected routing messages when the protection scheme is integrated with the routing protocol. In ad hoc mode, IEEE 802.11i requires a pre-shared symmetric group key. Unique identification of the source is precluded, and IEEE 802.11i only provides hop-by-hop protection.

Routing messages are often point-to-multipoint transmissions. IPsec is geared towards two communicating parties that establish a secure channel with the aid of a unique symmetric key. That is, IPsec also allows protection of multicast traffic based on pre-shared group keys. The security offered is then parallel to IEEE802.11i. The routing messages will be protected hop-by-hop, and not end-to-end. Intermediate nodes in possession of the group key can undetectably modify the relayed routing messages. Protection against insiders exhibiting Byzantine behavior is easier to achieve with an asymmetric component. An RFC describing the use of RSA-based digital signatures with ESP and AH has recently been published [104].

Another challenge is that IPsec and IEEE 802.11i protects on a per-routing packet/frame basis, but the granularity needed may be on a per-message basis. Each packet may contain more messages, possibly from different sources.

A generally adopted assumption is that ad hoc networks are IP based. The ad hoc mode of IEEE 802.11 has become the most important MAC protocol for MANETs. But other protocols may also be used. Relying on the lower layers for protection of routing information places restrictions on the choice of protocols.

Altogether, end-to-end unilateral authentication of routing information implies modification of the existing IPsec and IEEE 802.11i standards.
2.2.8 Summary

Due to the wireless media and ad hoc nodes operating both as routers and communication end-points, the routing protocols of ad hoc networks are more prone to attacks than routing protocols in fixed networks. Possible attacks comprise a number of attacks that all can be classified as various forms of denial of service attacks.

Whereas the secure ad hoc routing protocols assume the protection is an integral part of the routing protocol, the ad hoc routing protocols specified in RFCs and Internet-Drafts from the IETF MANET WG and the IETF OSPF WG assume security is enforced with the aid of lower layers security mechanisms such as IPsec AH or IEEE 802.11i. However, due to the point-to-multipoint transmission and the need for each receiver to be able to validate the authenticity of the routing messages; modifications in the existing standard may be required.

2.3 Key Management

The security and the applicability of any cryptographic scheme depend upon a proper key management scheme. Key management comprises key generation, validation, storage, distribution, update, revocation and deletion. Furthermore, key lengths, the frequency of key changes/updates and how keys are used play an important role for the security of the scheme.

Key management schemes proposed for ad hoc networks surveyed in paper I range from self-organizing contributory schemes to managed distributive ones. The proposals include both symmetric, identity-based public key as well as traditional certificate-based public key schemes. The next sections present some important aspects of key management and state-of-the-art that are not addressed or less emphasized in the articles presented in part II.

2.3.1 Key generation and authentication

No matter whether cryptographic keys are generated and distributed by a single entity or set up through a collaborative effort of multiple entities, mutual authentication of involved parties is required in order to provide a secure key management system.
Secure initialization of security associations can be achieved through physical proximity and out-of-band identification. Methods proposed so far for ad hoc networks rely on demonstrative identification through physical contact or range-limited channels [7] [97] [98] or trust evaluation through own observations [101]. The resurrecting duckling model presented in [97] and [98] assumes that the first neighbor a node (the duckling) sees, is its trustworthy mother. It therefore accepts imprinting of keys from this node. The imprinting means a key is transferred in plaintext from the mother node. It is not evident what should happen if the mother duck does not want its duckling.

SUCV [67] and CAM [96] suggest identifiers derived from public keys of private/public key pairs. The aim is to provide unique and cryptographically verifiable IPv6 addresses. There are no certificates and trusted third parties. The nodes simply generate a private/public key pair, and derive the address from the public key. Only the node that knows the secret public key can provide a signature that can be verified with the public key. The technique prevents that other nodes can easily masquerade as one of the other nodes. But the method cannot be used to distinguish trustworthy from untrustworthy nodes, as assumed in the Simple Ad hoc Key Management protocol (SAKM) [111]. Nor does the method prevent Sybil attacks.

2.3.2 Key Lengths and security levels

The key lengths must be chosen large enough to provide an adequate level of security. Table 2 shows hash lengths and key sizes of comparable security levels in accordance with the recommendations in [57], [58] and [75]. The security level is measured in bits of a symmetric key.

Table 2 shows that the required key lengths and asymmetric signature sizes are expected to increase over time. It also shows that the number of bits in traditional asymmetric schemes grows faster than the hash sizes and the symmetric key lengths. Contrasting hash functions and symmetric schemes, the computational cost of asymmetric schemes generally increases more rapidly with augmenting key lengths.

The estimates in Table 2 assume that whereas more efficient attack methods are found for asymmetric schemes, exhaustive key search is the most efficient method available for breaking a symmetric cryptosystem. The security of asymmetric schemes rests on different computational problems such as the difficulty of factoring large
integers and the discrete log problem. The complexity of solving these problems depends on the best factoring and discrete log algorithms available. The estimates in \[57\] and \[58\] are based on the assumptions that cryptanalytic progresses follow the same trend in next decades as in the past decades. It is assumed that a cryptosystem offers adequate protection until a given year if the cost of a successful attack measured in that year can be expected to be about 40M dollardays, \textit{i.e.}, with 40 million dollars it would take one day to break. Anyone who believes that the assumptions do not hold should adjust the key length accordingly.

MANET routing information has only instant value. Ad hoc networks are expectedly short term networks formed for a specific event. The lifetime may range from hours to days, or possibly weeks. A key that is disclosed when the network terminates is of little value to an attacker. A malicious node in possession of a valid key may only harm the routing as long as the compromise is not detected and the key revoked. Even if an attacker succeeds in occasional signature forgeries the network service may not be severely affected. With sequence numbered and periodical routing messages, the attacker must be able to provide more forged signatures in sequence in order to succeed in a persisting attack.

**Table 2 Comparable security levels**

<table>
<thead>
<tr>
<th>Source</th>
<th>Security level (bits)*</th>
<th>Hash Size**</th>
<th>IFC &amp; TDL N</th>
<th>FFC (SDL) $p/q$</th>
<th>ECC key length</th>
<th>Safe until year</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIST [75]</td>
<td>80</td>
<td>80</td>
<td>1024</td>
<td>1024/160</td>
<td>160-223</td>
<td>2010</td>
</tr>
<tr>
<td></td>
<td>112</td>
<td>112</td>
<td>2048</td>
<td>2048/224</td>
<td>224-255</td>
<td>2030</td>
</tr>
<tr>
<td>Lenstra [57]</td>
<td>80</td>
<td>80</td>
<td>$\geq 1280$</td>
<td>$\geq 1280/160$</td>
<td>160</td>
<td>2018</td>
</tr>
<tr>
<td></td>
<td>112</td>
<td>112</td>
<td>$\geq 3072$</td>
<td>$\geq 3072/223$</td>
<td>224</td>
<td>2065</td>
</tr>
<tr>
<td>Lenstra &amp; Verheul [58]</td>
<td>80</td>
<td>80</td>
<td>1464</td>
<td>1120/141</td>
<td>149-165</td>
<td>2012</td>
</tr>
</tbody>
</table>

*Security level is measured in bits of a symmetric key, **Hash size refers to preimage resistance (collision resistance: double the number of bits)*

These factors indicate that shorter key lengths may suffice for ad hoc routing information. Short keys and small signature sizes are also beneficial for bandwidth efficiency. However, factors such as the expected key period may pull in the opposite direction. Key changes are likely to represent a significant administrative cost, especially in organizations with a large number of nodes. Keys can therefore be expected to last for more than one MANET instantiation. In addition, frequent changes of keys that are used to bootstrap the network service are undesirable as they represent a threat to network availability. A framework for the decision of proper key lengths in ad hoc networks remains a topic for further work.

2.3.3 Revocation and exclusion of nodes

The exclusion of a single node in systems relying on a symmetric group key implies a re-keying process where the remaining nodes switch to a new group key not known by the excluded node. A typical solution includes a key distribution centre that sends the new group key encrypted with unique keys for each of the remaining nodes. Connectivity to a key distribution centre at all times cannot be guaranteed in a mobile ad hoc network. Furthermore, distribution of a new group key imply a bandwidth consuming process, and there is no guarantee all nodes receive the new key in a timely manner. As the group key is used to bootstrap the protected network service, the re-keying represents a threat to network availability.

An alternative to re-keying the remaining nodes is to use key-zeroing messages that delete the key material in the node to be excluded, e.g., as proposed for VANETs in [45]. Still, this approach is better suited for infrastructure based networks with guaranteed connectivity and where the nodes have to “sign in” at a base station or access point to obtain access to the network. The solution requires access to a centralized key management centre and tamper-proof devices (otherwise the group key could be extracted before erasure and exploited by malicious nodes). By default, there is no guarantee that the node to be excluded actually receives the key-zeroing message. Hence, the key erasure should be acknowledged by the excluded node. Furthermore, the management centre must be able to verify the integrity and authenticity of origin of the acknowledgement. This typically requires the management centre shares a unique symmetric key with each node. Furthermore, there is no guarantee that the management centre actually receives the acknowledgement on
the first attempt. Although still challenging, revocation is easier with asymmetric schemes.

Several of the proposals from the OSPF WG and the MANET WG assume protection is added by IPsec or IEEE802.11i. Schemes for key changes or revocation of keys to exclude nodes during MANET operation are not part of these standards. With IEEE802.11i in infrastructure mode of operation, the nodes can be denied access when they try to reconnect to the wireless network. In ad hoc mode of operation there is no such possibility. Standard key management protocols for IPsec such as IKEv2 [47] does not apply to point-to-multipoint traffic. Nor does IPsec specify methods for revocation or changes of keys used to protect multicast traffic. In the secure multicast architecture described in RFCs 3740 [30], a central group controller / key server administers the security associations.

Revocation is to a little extent addressed in the secure ad hoc routing proposals described in section 2.2.5. ARAN [94] assumes a trusted certificate server broadcasts a signed revocation message informing which certificate has been revoked.

Revocation schemes proposed for ad hoc networks typically rely on accusations and the exchange of revocation information between the ad hoc nodes, plus downloads from a trusted entity during periods with connection to the fixed network [20] [21] [45] [68]. Surveys of revocation schemes for fixed net are found in [108] and [112].

The detection of compromised nodes and the decision to revoke a key can be made on the background of reports of stolen items, monitoring of node behavior, accusations of malicious actions or other intrusion detection schemes. Intrusion detection schemes for mobile ad hoc networks are surveyed in [3] and [66].

2.3.4 Summary

A proper key management system is the basis for any cryptographic authentication schemes. The initial key generation and distribution requires some form of physical or out-of-band authentication of the involved parties. The short term value of ad hoc routing information indicates shorter key lengths than for sensitive payload. However, the key periods may exceed the lifetime of an ad hoc network instantiation. The keys may be (re)used in more MANETs. This pulls in the direction of increased key sizes.
A framework for the decision of proper key lengths for ad hoc routing information is a topic for further research. Revocation is also not a fully solved problem.
3 Contributions of the thesis

The contribution of this thesis is mainly within authentication of routing messages and key management. These are key elements for securing MANETs for emergency and rescue operations. Appropriate security mechanisms are proposed and evaluated.

The majority of secure ad hoc routing protocols proposed so far, tend to focus on the protection techniques, rather than the computational cost and bandwidth consumption. Our work highlights the importance of taking these factors into consideration in the design of secure routing protocols. In addition, attention must be paid to the nature of routing control traffic and its discrepancies from ordinary application data.

Figure 4 outlines the contributions and the flow of the work conducted. The survey of key management schemes [34] showed that no single technique dominates. It revealed in particular that bandwidth consumption has been neglected. This is an important aspect in ad hoc networks. An interesting alternative is ID-based systems that make certificates superfluous. This initiated the analysis of ID-based signatures for protection of ad hoc routing information in paper II [33].

Figure 4 Outline of the work and contributions
The ability to scale remained an issue even though ID-based systems remove the bandwidth cost caused by certificate exchanges. We therefore proposed a hybrid solution. The hybrid protection scheme for OLSR (proposed in paper III) [32] reduces both the overhead and computational cost.

In the last paper [31] we looked at revocation. A proper revocation scheme enables exclusion of compromised and malfunctioning nodes. We proposed a simple and robust, yet bandwidth efficient method for the distribution of revocation information in ad hoc networks. The next sections summarize the contributions of each article in more detail.

3.1 Contribution of Paper I: A Survey of Key Management in Ad Hoc Networks

Authentication is a fundamental issue for a reliable network service. The nodes must be able to distinguish trustworthy from untrustworthy nodes in the neighbor discovery process, and they must be able to verify both routing message origin and integrity.

Any cryptographic authentication scheme requires proper key management. That is, mutual authentication of the involved parties is required during the key setup. This is typically solved through physical proximity and some out-of-band proof of identity. Ad hoc networks with a planned origin rooted in organizational structures where pre-configuration is possible, makes the key setup easier than in truly ad hoc environments where the nodes have no prior trust relationships. Still, it may be desirable to include nodes from different security domains, e.g., in cross-border disaster relief operations.

This paper provides a comprehensive overview of key management schemes for ad hoc networks. It also presents an evaluation of their applicability for initialization and maintenance of a protected network service.

The key management schemes are classified as either contributory or distributive, and are evaluated according to a set of evaluation criteria. The criteria cover applicability, security, robustness, scalability and simplicity. Applicability refers to fundamental assumptions such as network origin, network size, mobility, range and human involvement. The applicability in ad hoc networks with nodes from a single security domain and multiple security domains is evaluated. Security includes metrics such as authentication, intrusion tolerance, trust management, revocation and...
vulnerability. The key management system should survive despite denial-of-service and be robust to Byzantine behavior. Timeliness and strict synchronization are other factors. The key management scheme should be simple and scale well enough to handle the expected number of nodes.

None of the studied key management schemes were found superior to one another in all situations. Several of the proposals made assumptions that do not fit the nature of ad hoc routing protocols and the requirements for network availability. Examples include assumptions such as guaranteed connectivity during key setup, a common view upon node ordering, and re-keying every time a node leaves or joins the network.

In general, the proposed schemes did not state for what purposes the keys were intended. Keys for protection of application data were typically implicitly assumed. But the pre-conditions for ordinary payload and network layer routing information differ. The setup of keys for ordinary application data can assume an already running network service. Key setup for routing information cannot.

The performance analysis tends to be restricted to computational complexity, whereas the practical constraint of MANETs may equally well be communication capacity. Both the size and number of messages are important. Due to the overhead introduced by lower layers of the communication protocol stack, the number of messages can have more impact than the size.

The evaluation revealed a lack of appropriate, robust and bandwidth efficient key management solutions.

3.2 Contribution of Paper II: Analysis of IBS for MANET Security in Emergency and rescue Operations

A reliable network service can be achieved with the aid of authenticated routing messages. If identifiers that are by default sent as part of the routing messages can be used as public keys, identity-based schemes may potentially scale better than traditional certificate based digital signatures. This paper analyses the consequences of IBS protected ad hoc routing messages. The optimized link state routing protocol is used to exemplify the implications of IBS protected network services.

Identity-based signature (IBS) schemes necessitate long term identifiers. Otherwise, the key administration cost becomes excessive. Keys used to protect the network layer should be related to the identifiers known by the routing protocol. In
order to provide unique IP addresses, a number of schemes for MANET address assignment and how to detect and resolve duplicate addresses have been proposed. Some suggestions are described in [1], [15], [81], [82], [100], and [102]. A survey of proposed methods is found in [8].

The requirement for long term identifiers restricts the possibilities for dynamic address allocation. However, the proposed solutions generally assume that any node is welcomed in the network. This is not necessarily the case in dedicated networks for emergency and rescue operations. The structure of emergency and rescue operations may enable address assignment co-located with the key administration. Fixed addresses may be acceptable. With IPv4, the MANET may be interconnection to the Internet through Mobile IP [84].

Asymmetric cryptographic schemes are known to introduce a considerable amount of overhead and resource consumption. Our analysis revealed bandwidth cost as a more limiting factor for the applicability of the signature scheme than the processing cost. The channel may become congested before the processing capabilities are exhausted. The calculations showed that signatures sizes of more than a few hundred bits are only applicable in small ad hoc networks.

The calculations assumed that the routing protocol control traffic was sent at a rate of 1Mbps utilizing IEEE 802.11b for the lower layers. However, other routing protocols, higher data rates or other MAC layer protocols could have uncovered the processing cost as the most critical factor. In any case, it shows the importance of considering both channel occupation as well as computational cost. With resource constrained nodes, little overhead as well as low computational cost is beneficial.

3.3 Contribution of Paper III: Hybrid Protection of OLSR

Symmetric schemes are efficient both regarding computational efficiency and bandwidth consumption. However, with a group key, the security is compromised with a single captured node. Source authentication is precluded, and malicious insiders may masquerade as others. Furthermore, exclusion of compromised nodes is challenging. The alternative of pairwise exclusive keys provides better intrusion tolerance, but scales badly.

Digital signatures and IBS schemes provide intrusion tolerance, and in addition enable unique identification and authentication of the origin of routing messages.
Another advantage is that revocation of asymmetric keys enables exclusion of compromised and malicious nodes without other nodes having to change their key. These are desirable features in the protection of ad hoc routing information.

Asymmetric schemes are known to be computationally expensive. Another drawback of traditional digital signature schemes is the potential need for bandwidth consuming certificate exchanges. The identity-based signature (IBS) schemes make certificates obsolete.

This paper proposes a hybrid protection scheme for proactive link-state routing protocols, and provides specific formats for its implementation with the optimized link state routing protocol (OLSR). The hybrid protection scheme suggests asymmetric, identity-based signatures combined with shorter and computationally cheaper hash values from a hash chain. The aim is a bandwidth efficient solution providing an adequate level of security. Both the security and the performance of the protection scheme are analyzed.

Bandwidth and computational cost is saved by signing only some routing messages. Unsigned messages include a hitherto undisclosed value from the hash chain. The hash values serve as lightweight proofs of authenticity, and make successful attacks harder. The hash values also simplify the detection of message sequence number wrap-around and make replay attacks easier to detect. The hash values prevent both inside attackers as well externals from successfully inserting unscheduled routing messages on the behalf of other nodes. However, under specific circumstances, as discussed in the security analysis, there is a risk that malicious nodes may succeed in temporarily introducing false routing information. An informal security analysis shows that attacks are not straightforward to mount. Signed routing messages will erase any false information.

Whereas the hybrid protection scheme is proposed for OLSR, the method may also be used for other proactive link-state routing protocols. Furthermore, the IBS scheme could be replaced with a digital signature scheme. Certificate distribution must then be taken into consideration.

The performance evaluation in the paper assumes asymmetric schemes and hash functions of similar security levels. It could be argued that more trust is put in the signed message than the ones followed by a hash. This implies a shorter hash value may be sufficient. However, the hash values represent a small overhead compared to the signatures. Reducing (or increasing) the hash length a few bits or bytes has little
impact on the bandwidth efficiency. For practical reasons both signatures and hash values should fit with byte boundaries.

3.4 Contribution of Paper IV: On the Distribution of Revocation Information in Ad Hoc Networks

Most practical communication systems will require a way to expel compromised, malfunctioning or maliciously behaving nodes. An appropriate revocation scheme is required.

The revocation methods for fixed networks are generally not very well suited for the ad hoc environment. Some require guaranteed connectivity to a central trusted entity to check certificate status on-line. Others trade update cost for timeliness or vice versa. The online certificate status protocol (OCSP) [70] and the Certificate Revocation Status (CRS) [65] are representatives of the first. The well known certificate revocation list (CRL) method is one example of the latter.

The CRLs usually include all revocations within the security domain, and are typically updated weekly or every second week [112]. Keys used to protect the network service of a MANET that lasts hours or days, necessitate higher granularity.

Including only certificates that have been revoked since last CRL update, $\Delta$-CRLs are shorter and provide fresher information. However, $\Delta$-CRLs are also normally issued with a frequency, e.g., daily, that may be to low. Higher frequencies lead to a significant update cost. Besides, if one or more of the ad hoc nodes did not receive the last CRL, the CRL may need to be distributed in the MANET anyway. Moreover, CRLs and $\Delta$-CRLs normally list revocations concerning the entire security domain and not only the subset of nodes that takes part in a specific MANET. This represents a waste of bandwidth.

There may be a large number of nodes in the security domain corresponding to the national emergency and rescue organizations. But only a subset of the nodes will normally take part in a specific operation and the MANET formed for this event. If the revocation information concerns other keys than those used by the nodes in the network, there is no need to distribute this information over the ad hoc network.

Schemes proposed for ad hoc networks typically rely on the ad hoc nodes’ exchanges of accusations or CRLs and downloads of CRLs from trusted entities during periods with connection to the fixed net. Accusations may lead to revocations
on false reasons. The MRL (MANET revocation list) scheme introduced in this paper ensures that only revocation information relevant for the particular MANET is distributed into the ad hoc network.

The MANET is assumed to be connected to the external network through one or more trusted gateways. The trusted gateway periodically informs the central trusted entity about the continued existence of the MANET and reports any new nodes detected. In response, the trusted entity returns a MRL listing revocations (if any) concerning keys of the (accumulated) nodes reported. Only non-empty MRLs are flooded into the network from the gateway.

The trusted gateways must be able to detect which nodes are in the MANET. This comes intrinsically with proactive ad hoc routing protocols, and may be achieved at some additional cost with reactive protocols.

The scheme is intended for ad hoc networks with a planned origin, and where a common point of trust exists. It is designed for revocation of keys used to protect the ad hoc network service, and the MRLs are distributed as an integral part of the routing protocol. The trusted entities are assumed to reside in the external network, but could also reside inside the MANET. Robustness can be strengthened through redundant trusted gateways.

The method can be used both for certificate revocation as well as revocation in identity-based public key schemes. The complete revocation lists are assumed to be distributed in a preparation phase prior to ad hoc network participation. But there is no need for the revocation list held by each participating party to be synchronized. And there is no need for the trusted gateways to synchronize the MRLs they spread into the network. On this background, the MRL scheme provides a simple, scalable, efficient and robust solution for the distribution of revocation information in ad hoc networks with a planned origin where common points of trust exists.

The article shows that there is a gain with the MRL scheme compared to CRL and $\Delta$-CRLs even if the revocation information is inserted through only one gateway. With more gateways pouring the revocation information into the network, the bandwidth savings obtained with the MRL scheme compared to the other two methods becomes even more significant. The MRL technique also scales well with multiple security domains involved. Contrasting CRL and $\Delta$-CRLs the size and frequency of revocation messages depends upon the number of nodes in the ad hoc network rather than the number and size of the involved security domains.
4 Concluding Remarks

Security is important in any communication system for emergency and rescue operations. We have concentrated on a subset of the aspects of security in ad hoc networks for such applications; protection of the routing protocols. Our research indicates that a reliable ad hoc network service is achievable. From this perspective there is nothing that precludes the use of MANETs in emergency and rescue operations. Related work, such as [105], which considers other aspects of ad hoc networks in emergency and rescue operations, points in the same direction. This indicates that ad hoc networks can become an important component in future disaster relief networks. And it is important that they are included in plans for future emergency and rescue communication systems.
References


PAPER I: A Survey of Key Management in Ad Hoc Networks

Anne Marie Hegland, Eli Winjum, Stig F. Mjølsnes, Chunming Rong, Øivind Kure and Pål Spilling

IEEE Communications Surveys & Tutorials, Vol. 8, no. 3, 3rd Quarter, 2006, pp. 48-66
A Survey of Key Management in Ad Hoc Networks

Anne Marie Hegland*, Eli Winjum*, Stig F. Mjølsnes†, Chunming Rong§, Øivind Kure† and Pål Spilling*

*UniK, Kjeller, Norway, §UiS, Stavanger, Norway, †NTNU, Trondheim, Norway
{amh;ewi;okure;paal}@unik.no; chunming.rong@uis.no

Abstract

The wireless and dynamic nature of mobile ad hoc networks (MANETs) leaves them more vulnerable to security attacks than their wired counterparts. The nodes act both as routers and communication end-points. This makes the network layer more prone to security attacks. A main challenge is to judge whether or not a routing message originates from a trustworthy node. The solution thus far is cryptographically signed messages. The general assumption is that nodes in possession of a valid secret key can be trusted. Consequently, a secure and efficient key-management scheme is crucial. Keys are also required for protection of application data. However, the focus here is on network-layer management information. Whereas key management schemes for the upper layers can assume an already running network service, schemes for the protection of the network layer cannot. Keys are a prerequisite to bootstrap a protected network service.

This article surveys the state of the art within key management for ad hoc networks, and analyzes their applicability for network-layer security. The analysis puts some emphasis on their applicability in scenarios such as emergency and rescue operations, as this work was initiated by a study of security in MANETs for emergency and rescue operations.
I INTRODUCTION

MOBILE ad hoc networks have wireless links and work independently of fixed infrastructure. They are self-organizing and self-configuring. The wireless nodes operate both as communication end-points as well as routers, enabling multi-hop wireless communication. The wireless devices imply limited power resources and bandwidth. Network topology may change rapidly due to mobility, interference, physical obstacles on the path, and so forth. Application areas range from conference hall networks to ad hoc networks for emergency and rescue operations and military tactical use.

The wireless and dynamic nature of ad hoc networks leaves them more vulnerable to security attacks than their wired counterparts. Passive eavesdropping as well as active message insertions, denial of service and battery-exhaustion attacks are inherently easy. Security attacks can be launched towards any layer of the protocol stack. Defense mechanisms for the lowest layers call for physical tamper protection and transmission security measures such as spread-spectrum techniques, frequency hopping, and interleaving. Cryptographic techniques are essential for the protection of the higher layers.

In wired networks the routers are part of an established and (at least to some extent) controllable infrastructure. The same is not true in ad hoc networks where the nodes act both as routers and communication end points. This makes the network layer more prone to security attacks. There is no guarantee that malicious nodes do not mingle and interfere. Examples of possible attacks are misdirection and insertion of bogus routing information, black holes (nodes attracting traffic by maliciously advertising shortest path to other nodes), and wormholes (adversary nodes colluding by tunneling packets from one part of the network to another).

A primary challenge is to decide which routing information can be trusted. A number of schemes relying on cryptographically signed routing messages have been designed – most without detailing key management further. Nevertheless, the possession of cryptographic keys serves as proof of trustworthiness. Consequently, a proper key-management service is required. This is to ensure that nodes which are legitimate members of the network – and only those – are equipped with the necessary keys whenever needed. Whereas key-management services are needed for application layer security as well as for protection of the network layer, this article focuses on the more challenging of the two, namely providing
keys for the network layer. Key management schemes for the application layer can assume an already running network service. Schemes for the network layer routing information cannot. Keys are a prerequisite to a protected network service.

This survey was motivated by an investigation of security in ad hoc networks for emergency and rescue operations. Most of the discussions have relevance for mobile ad hoc networks (MANETs) in general. However, emergency and rescue operations have some additional requirements. Where appropriate, concerns regarding the applicability of the various key-management schemes in settings with the characteristics of emergency and rescue operations are highlighted.

Emergency and rescue operations imply MANETs with nodes that have gone through a preparation phase prior to MANET initialization. That is, predistribution of keys and other parameters is possible. MANETs for emergency and rescue operations present stronger requirements than most commercial applications. Time is scarce. When the rescue team arrives on the scene of an accident, communication should be established immediately and maintained with as little human interaction as possible. Availability is a number one requirement.

The network resources should be reserved for the members of the emergency and rescue team, and not used to convey arbitrary data for others. It should be possible to distinguish legitimate nodes from untrustworthy ones and build a reliable route through trusted nodes only.

The structure of emergency and rescue operations has implications for key management as follows.

--- Single administrative domain involved (SAD) ---

SAD operations refer to operations where all involved parties belong to the same regime or share a common, predefined point of trust. Local, regional or national rescue operations including only predefined actors are in this category. This setting enables pre-configuring of security credentials.

--- Multiple administrative domains involved (MAD) ---

MAD operations represent operations involving ad hoc partners. That is parties that have had no prior contact and belong to different organizational/security domains. This means cases where overall preconfiguring of security parameters is not possible. Examples include
cross-border operations and operations involving industrial companies.

Standards: None of the emerging MANET Internet-drafts and RFCs have so far encompassed key management. Of other standards, the IEEE 802.11i [1] security amendment for IEEE 802.11 wireless local area networks assumes keys are preshared or established with the aid of fixed infrastructure. In case of truly ad hoc communication, preshared symmetric keys are the only option. The aim of IEEE 802.11i is protection of payload (data frames) on layer 2. IEEE has in 2005 begun work on 802.11w that will cover security on management frames. Other standards for wireless communication include the ZigBee [2]/IEEE 802.15.4 [3] and the Bluetooth [4] specifications for personal area networks. The preconditions of these standards are infrastructure-based networks and do not apply to MANETs. ZigBee specifies key management for the security elements of IEEE 802.15.4. ZigBee assumes the initial keys are predistributed, installed out-of-band, or received in the clear over the air from a trust center. Keys in Bluetooth are derived with the aid of PIN codes. A common PIN code is entered out of band in pairs of nodes that wish to communicate. Standards for key management in ad hoc networks lack.

The contribution of this paper is a survey of proposed key management schemes for ad hoc networks, and an analysis of their applicability for MANET network layer security – with some emphasis on their applicability in MANETs for emergency and rescue operations.

The rest of the paper is organized as follows. Desirable features for key management schemes in MANETs are described. Next we classify and evaluate proposed key-management schemes. Conclusions are then made.

II Desirable Features of MANET Key Management Schemes

The evaluation parameters reflect the intention of bootstrapping a protected MANET network service. Evaluations of key-management protocols often focus on computational complexity. However, with the wireless media, bandwidth is regarded as a more constrained resource than computational power. A key-management service for protection
of the network layer should not demand an already running network service or overall connectivity.

—**Applicability:** The various key-management schemes focus on different targets. The aim may range from group key establishment to availability of central management entities. Their applicability for SAD and MAD operations depends on the fundamental assumptions as to network origin (planned or truly ad hoc), network size, node mobility, geographic range and the required level of human involvement.

—**Security:** *Authentication* and *intrusion tolerance* is a primary concern to ensure no unauthorized node receives key material that can later be used to prove status as a legitimate member of the network. Nobody should provide private keys or issue certificates for others unless the others have been authenticated. Intrusion tolerance means system security should not succumb to a single, or a few, compromised nodes. Other central security issues are *trust management* and *vulnerability*. Trust relations may change during network lifetime. The system should enable exclusion of compromised nodes. In order to judge the security of a key-management scheme, possible vulnerabilities should be pinpointed. Proper key lengths and cryptographic algorithms of adequate strength are assumed.

—**Robustness:** The key-management system should survive despite denial of service attacks and unavailable nodes. The key-management operations should be able to be completed despite faulty nodes and nodes exhibiting Byzantine behavior, that is, nodes that deliberately deviate from the protocol. Necessary key-management operations caused by dynamic group changes should execute in a timely manner. Key management operations should not require networkwide and strict synchronization.

—**Scalability:** Key management operations should finish in a timely manner despite a varying number of nodes and node densities. The fraction of the available bandwidth occupied by network management traffic should be kept as low as possible. Any increase in management traffic reduces available bandwidth for payload data accordingly. Hence, scalability of key-management protocols is crucial.

—**Simplicity:** Simplicity regarding user-friendliness and communication overhead is an additional intuitive and overall critical factor to the success of a key-management scheme. We reckon, however, that a system that is secure, robust and scalable implies simplicity.
Given that these conditions are fulfilled, we believe simplicity is first and foremost a matter of implementation.

The ideal key-management service for ad hoc networks should be simple, formed on the fly, never expose or distribute key material to unauthorized nodes, ensure that system security does not succumb to (a few) compromised nodes, easily allow rekeying/key updates, enable withdrawal of keys when nodes are compromised or keys for other reasons should be revoked, be robust to Byzantine behavior and faulty nodes, scale well enough to handle the expected network sizes and node densities, and efficiently manage network splits and joins.

Signed routing information requires a security relation that allows one-to-many signing and verification. Routing messages are often broadcast, and all receiving nodes should be able to check the validity. Messages such as neighbor-detection messages are not forwarded by other nodes. Other routing messages, such as topology-information messages in proactive routing protocols and route requests and route replies in reactive routing protocols, are flooded into the entire network. The receiving nodes may not be known to the transmitting node. In addition, bandwidth is limited. Unique signatures for each receiver scale badly. In other words, pairwise keys provide no good option for protection of routing information.

III Overview of Key Management Schemes

Key management schemes can be classified in several ways. In this article we have the main categories contributory and distributive, rather than the more commonly used contributory and centralized. The distributive category is here defined to encompass schemes where each key originates from a single node. The nodes may very well cooperate during key distribution, but any key originates from a single source.

Distributive schemes may be centralized, but can also be distributed. In the latter, each node generates a key and tries to distribute it to others. In contributory schemes, the key is a result of a collaborative effort of more nodes. Some of the contributory schemes studied here rely on a centralized entity, others do not. Altogether, we chose the categories “contributory” and “distributive” as we found this classification best reflected the origin of the keys in the schemes studied here. Our classification is illustrated in Fig.1.
The distributive category is divided into symmetric and public key schemes. Public key schemes include traditional certificate-based and identity-based schemes. The symmetric schemes are classified as either MANET schemes or WSN (wireless sensor networks) schemes. WSNs represent a new class of ad hoc networks with more constrained nodes than traditional MANETs.

**Contributory** schemes are characterized by the lack of a trusted third party responsible for generation and distribution of the cryptographic keys. Instead, all communicating parties cooperate to establish (i.e., “agree” upon) a secret symmetric key. The number of participants ranges from two parties (establishing a pairwise key) to many parties (establishing a group key). Although not necessarily designed with ad hoc networks in mind, intuitively the contributory approach of collaboration and self-organization may seem to fit the nature of ad hoc networks. A number of contributory schemes are therefore reviewed and evaluated later. Only one of these was designed specifically for ad hoc networks.

**Distributive** schemes involve one or more trusted entities and comprise both public key systems and symmetric systems. Truly ad hoc networks require the trusted entity to be established impromptu during network initialization.

**Certificate-based public key schemes** require the public keys to be distributed in a way that allows the receiving nodes to verify the authenticity of the key material. The wired network solution is a public key infrastructure (PKI) where a centralized certificate authority (CA) issues certificates binding the public keys to specific users/nodes.

If it is suspected that a node has fallen into the wrong hands, or the node for other reasons should be expelled, the certificate is revoked. Revoked certificates are added to the
certificate revocation list (CRL). The CA signature guarantees the authenticity of certificates and CRLs. Under the assumption that a centralized trusted entity is not well suited for ad hoc networks where overall availability cannot be guaranteed all the time, the proposed key-management schemes for ad hoc networks involving certificate-based PKI, advocate various ways to distribute the CA functionality. The intuitive approach of naive CA replication is not reckoned as good enough as it poses poor intrusion tolerance. With more nodes holding the private CA key, the higher is the risk of getting it compromised.

Identity-based public key schemes [5] represent a new type of public key system. They allow user identities (e.g., e-mail - or IP addresses) to be used as public keys, and make certificates superfluous. A trusted entity is however required in order to generate and distribute the private keys corresponding to the various identities. The trusted entity is also needed for revocation. The trusted entity may sign a list of withdrawn identities. As with traditional public key systems, it has been suggested to spread the trusted entity over more nodes.

Symmetric systems aim at distributing one or more shared secrets through secure channels. Many of the symmetric key management systems for ad hoc networks found in the literature are intended for Wireless Sensor Networks (WSNs). The sensor nodes possess very limited power, memory and computational resources compared to traditional MANET nodes. Symmetric systems may thus be the only option. WSNs normally include a base station. That is, WSNs have a certain amount of infrastructure and are thus not truly ad hoc networks. This survey distinguishes between symmetric schemes for traditional MANETs and WSN schemes. A number of WSN schemes have been included in order to evaluate their applicability in traditional MANETs.

Related surveys of key management-schemes: A survey of key distribution mechanisms for wireless sensor networks is found in [6]. Key Management schemes for secure group communication are surveyed in [7]. Reviews of key management protocols for ad hoc networks and sensor networks are also found in [8-10]and [11].

A. Contributory Schemes

The main implications and limitations of various types of contributory schemes in ad hoc networks are demonstrated by the schemes studied in this section. Table 1 summarizes the features of the different schemes.
**Figure 2.** Outline of the contributory schemes (all exponentiation of generator g are modulo prime p)
1) **Diffie-Hellman (D-H)** – D-H [12] establishes a unique symmetric key between two parties. It relies on the discrete log problem (DL); deciding $S$ given $g^S \mod p$ being a hard problem. D-H is outlined in Fig.2. The parties agree upon a large prime, $p$, and a generator, $g$. Each party randomly chooses a secret, $S_A$ and $S_B$, and transmits the public values, $(g^{S_A}) \mod p$ and $(g^{S_B}) \mod p$, as shown in the figure. Raising the number received from the other party to the power of its own secret, gives a common secret key, $g^{(S_A S_B)} \mod p$, shared only by the two.

Like any schemes involving pairwise unique keys, D-H provides intrusion tolerance. A captured node only compromises the keys it shares with its communicating peers. Byzantine and faulty nodes basically only disturb their own key establishment with communicating peers. D-H is vulnerable to man-in-the middle (MIM) attacks. It is left for the nodes to judge who to trust. But as authentication lacks, Alice cannot be sure that she actually communicated with Bob and not Charlie.

The generic D-H scheme is not applicable for protection of routing information in ad hoc networks. It applies to two parties only. Protection of routing messages with pairwise keys necessitates a different signature for each possible recipient, which scales badly. D-H has been included in this survey solely because the majority of the contributory schemes are founded on this scheme. They basically seek to remedy the shortcomings of D-H regarding MIM vulnerability and extendibility to more than two parties.

2) **Ingemarsson, Tang and Wong (ING)** – ING [13] provides a symmetric group key by extending the two-participant D-H scheme to $n$ participants. Fig.2 shows the principles with 4 nodes. All nodes are arranged into a logical ring. After $n-1$ rounds, each node can calculate the secret key. Each round involves an exponentiation from every node, and every node must transmit its share to the next node in the logical ring as shown in the figure.

ING lacks authentication and is vulnerable to MIM attacks. It scales poorly. Communicational complexity grows proportionally to the number of nodes squared. Byzantine behavior or faulty nodes may inhibit successful key establishment. A captured node means the group key is compromised, and necessitates a rekeying. The scheme does not specify how compromised nodes can be detected. The requirement for the nodes to

---

1For simplicity, the key management schemes are given short names. The short names do not necessarily represent generally adopted abbreviations.
organize into a logical ring during the key agreement procedure makes ING unsuitable for ad hoc networks. The establishment of keys for protection of routing information implies a logical ring of one-hop neighbors only (all nodes within direct transmission range). With mobile nodes and unstable links it is questionable whether ING will ever complete successfully.

3) Burmester and Desmedt (B-D) – B-D[14] seeks to establish a group key. It relies on the DL problem. But, contrary to the other contributory schemes studied here, it is not based on D-H. An outline of B-D with four nodes is shown in Fig.2. B-D completes in three rounds. Every node picks a secret, $S_i$, and multicasts its public value, $Z_i = g^{S_i}$, to all other nodes in the group. In round 2, every node calculates and multicasts a new public value. This value is derived by dividing the public value received from the next node with the public value received from the previous node in the logic ring of nodes, and raising the result to the power of its own secret, $S_i$, as illustrated in the figure. In the third and final round every node calculates the conference key from its secret and the information received from all the other nodes in the previous rounds.

B-D is apparently more efficient than ING as it completes in three rounds. However, each round requires a high number of exponentiations and reliable multicast. Reliable multicasting is difficult in wired networks, and even more challenging in ad hoc networks. Changes in group membership necessitate a restart of the key-agreement procedure. In an ad hoc network with moving nodes it may thus never be possible to establish a group key by B-D, nor handle later changes in group membership. Group changes will certainly cause delay and disruption. B-D also demands an already running routing protocol or only one-hop neighbors, that is, the key-agreement schemes depend on an already established routing infrastructure – but the infrastructure cannot be established before the keys have been set up. B-D authentication of the public values (not shown in the figure) can be implemented with the aid of pre-distributed public keys. Trust is managed through the certificate issuer. This implies a planned network and the basic key-management problem reverts to a public key scheme.

4) Hypercube and Octopus (H&O) – H&O [15] reduces the number of rounds and exponentiations of ING from $n$ to $d$ ($n=2^d$) by arranging the nodes in a hypercube, that is, a $d$-dimensional cube. Figure 2 illustrates H&O in a network with 4 ($2^2$) nodes. In step 1,
nodes 1 and 2 perform a D-H key agreement. Nodes 3 and 4 do the same. The symmetric keys established in step 1 are used as the secret values in a new D-H key agreement in step 2: node 1 and 4 perform a D-H key agreement and node 2 and 3 do the same and so on. H&O actually consists of two protocols: Hypercube and Octopus. Hypercube (shown in the figure) assumes the number of participants is a power of 2. Octopus extends Hypercube to allow an arbitrary number of nodes.

H&O is vulnerable to MIM attacks as authentication is absent. Byzantine or faulty nodes may preclude successful key agreement. Changes in group membership require rekeying. It is left for the nodes to decide when re-keying is needed. Like B-D and ING, H&O relies on an underlying communication system to provide a consistent node-ordering view to all group members. Besides the difficulty of keeping a consistent node ordering where nodes join and leave dynamically, it implies an already running (unprotected) routing protocol or only 1-hop neighbors. The latter scales badly. Altogether, H&O is unsuitable for network layer security in ad hoc networks.

5) Password authenticated key agreement (A-G) – A-G [16] is the only one of the contributory systems studied that has been designed with ad hoc networks in mind [16]. A-G is basically H&O extended with password authentication as indicated in Fig.2. It assumes all legitimate participants receive a password offline (written on the conference hall blackboard or distributed through another location limited channel). The nodes must prove knowledge of the password during the pairwise D-H key agreements of the H&O protocols as shown in Fig.2. The figure shows the password authenticated key agreement between two nodes. The password is used to encrypt the public value and an initial challenge in a challenge-response protocol as illustrated in the figure.

A-G doubles the number of messages and increases the computational complexity compared to H&O. It remedies H&O’s vulnerability to MIM attacks at the price of scalability.

A-G inherits the deficiencies of H&O regarding dependability of an already established communication infrastructure and node ordering scheme. Hence, it is not suitable for network layer security in mobile ad hoc networks.

6) CLIQUES (CLIQ) - CLIQ [17, 18] is outlined in Fig.2. It extends the generic D-H protocol to support dynamic group operations. CLIQ distinguishes between initial key
agreement (IKA) and auxiliary key agreement (AKA). IKA takes place at group formation. AKA handles all subsequent key-agreement operations. In both cases, a group controller synchronizing the key agreement procedure is required.

The figure shows the IKA protocol with four nodes. Stage 1 (the upflow stage) starts from node 1 which picks a secret exponent, $S_1$, and unicasts $g^{S_1}$ to next node. Node 2 picks a secret exponent $S_2$, and unicasts to node 3 the values shown in the figure. The procedure is repeated until the final node – the group controller is reached. The group controller is now able to calculate the secret group key (i.e., the generator) raised to the power of the secret exponents of all nodes in the group. In stage 2 (the downflow stage) the group controller multicasts the intermediate values required by each of the other nodes to calculate the secret group key, as shown in the figure.

Both AKA (not shown in the figure) and IKA rely on the group controller. The group controller of CLIQ thus represents a single point of failure. Each AKA operation results in a new group key that is independent of all previous keys. Adding a new member with AKA basically extends stage 1 of the IKA protocol with one node. The role of the group controller can be fixed or floating. Allowing any node to take over the role as group controller renders the system vulnerable to malicious nodes. CLIQ omits authentication. The designers have left security properties (such as authentication) out while focusing on group changes, but argue that authentication could easily be added. Other major drawbacks with CLIQ, as with B-D, are dependency upon reliable multicast and availability of a consistent view of node ordering. With variable connectivity it is questionable whether IKA and AKA would ever complete successfully. With unstable links, highly mobile nodes and rapid splits and joins, instability may result.

7) Other contributory schemes –A large number of key agreement schemes relying on already distributed keys have been proposed. The basic key-management problem thus reverts to distribution of the initial keys. Several schemes are also two-party protocols unsuitable for network layer security and are therefore left out of further discussions. Examples include MQV[19] based on traditional public keys, schemes relying on identity-based public keys such as [20] and [21], and the D-H based protocols proposed in [22].

8) Summary of the contributory key management schemes: Although the contributory approach at first glance may seem to fit the self-organizing nature of ad hoc networks, none
of the contributory schemes are good candidates for key management in ad hoc networks. D-H, ING and H&O can be skipped due to missing authentication. They are vulnerable to MIM attacks. B-D and CLIQ can be left out – no matter whether the authentication scheme is included or not – as they have an inherent survivability problem with the dependency on reliable multicasting. A-G fails on scalability and robustness due to the dependency upon node ordering, and availability of all nodes during group changes.

B. Distributive Schemes

This section surveys public key and symmetric key management schemes proposed for ad hoc networks. Tables 2 and 3 summarize the features of the public key schemes and the symmetric schemes, respectively.

1) Public Key Schemes

--*Partially distributed Threshold CA Scheme (Z-H)* -Z-H[23] assumes a PKI system and puts forward a framework to provide an available, intrusion-tolerant, and robust CA functionality for ad hoc networks. The private CA key is distributed over a set of server nodes through a \((k,n)\) Secret sharing scheme [24]. The private CA key is shared between \(n\) nodes in such a way that at least \(k\) nodes must co-operate in order to reveal the key. (Finding the private CA key \(S\) is comparable to finding \(f(0)\) given a polynomial \(f(x)\) of degree \(k-1\) and knowing \(k\) values, e.g. \(f(1), f(2) \ldots f(k)\).)

When queried, each server generates a partial signature of the certificate using its private key share in a threshold signature scheme [25]. A server acting as combiner collects the partial signatures and produces a valid signed certificate.

Z-H advises *share refreshing* to counter mobile adversaries, i.e. adversaries that temporarily compromise one server and then attack the next. *Proactive secret sharing schemes* [26] allow the shareholders to periodically refresh their shares through collaboration. An adversary must thus compromise more than \(t\) shares *between* refreshes in order to compromise the system. The original secret does not change, only the shares held by the servers. Bear in mind the homomorphic property: If \((s_1,s_2,\ldots,s_n)\) is a \((k,n)\) sharing of \(S\) and \((a_1,a_2,\ldots,a_n)\) is a \((k,n)\) sharing of \(A\), then \((s_1+a_1, s_2+a_2,\ldots,s_n+a_n)\) is a secret sharing of \(S+A\) [27]. Choosing \(A=0\) gives a new sharing of \(S\). The scheme is made robust to missing
and erroneous shares through *verifiable secret sharing* [28]: Extra public information testifies to the correctness of each share without disclosing the share.

Although not clearly stated, the system relies on a central trusted *dealer* to bootstrap the key management service, and decide which nodes shall act as servers. Z-H assumes an underlying (unsecured) routing protocol.

According to [23] nodes cannot get the current public keys of other nodes or establish secure communication with others if the CA service is unavailable. However, every node should hold a copy of its own certificate. For network layer security, it would be more efficient to receive the certificates directly from the communicating peers (or other nodes in the neighborhood). If the certificate is needed to verify a signature on routing information, the node in question must certainly be available; otherwise there would be no requirement to verify its routing message. Thus, the need for on-line CA access is limited. Every node must contact the CA to get its initial certificates (and receive the public key of the CA). The same is true if the node for some reason has lost its private key or has had its certificate revoked. However, to get a new certificate, the node should be authenticated by the CA service – which necessitates some sort of physical contact between the node and the CA service. Certificate *updates* call for CA service. For scenarios like emergency and rescue operations, it would be better to make sure certificates are renewed in the preparation phase and not during network operation.

The CA service is needed for revocation and distribution of CRLs. Z-H postulates that public keys of nodes that are no longer trusted, or have left the network, should be revoked. In an ad hoc network it can be hard to decide when a node has actually left the network. Revoking keys due to temporal missing connectivity would not be wise. More important is revocation of keys belonging to captured nodes. The frequency of such revocations in networks for emergency and rescue operations will expectedly be low.

Periodical share refreshing implies some form of synchronization. Synchronization is bandwidth consuming and difficult in ad hoc networks. Management traffic between server nodes and certificate exchanges also consumes much bandwidth, and makes Z-H scale badly. A single CA or hierarchy of CAs is likely to prove better than the Z-H approach. SAD operations allow predistributed certificates. MAD operations call for on-scene cross-certification of the root CAs of the merging domains. Efficient spreading of the cross-
certificates in the respective domains is a problem for further investigation. There is no easy way to update the private/public CA key pair and make sure all nodes are informed.

---MOCA---MOCA [29,30] is basically an extension to Z-H [23]. The focus is on distributed CA services and communication between the nodes and the server nodes – MOBILE Certificate Authorities (MOCAs). Whereas Z-H does not state how to select CA servers, MOCA suggests the nodes that exhibit best physical security and computational resources should serve as MOCAs. The MOCA scheme furthermore “moves” the combiner function of Z-H from the CA servers to the requesting end-nodes. The benefit is a less vulnerable scheme as the nodes no longer depend on the availability of the CA server nodes to combine the partial certificate signatures.

A MOCA certification protocol, $MP$, is proposed to provide efficient and effective communication between clients and MOCAs. According to $MP$, certificate requests should be unicast to $\beta$ specific MOCAs that, based on fresh routing entries or short distance, are likely to be accessible. With the $(k,n)$ threshold scheme, $k$ MOCAs are required to complete a certification service. To increase probability of receiving at least $k$ responses: $\beta = k+\alpha$. When availability drops, the protocol returns to flooding (as in Z-H). It is assumed that MP maintains its own routing tables and co-exists with a “standard” ad hoc routing protocol.

Placing the CA servers in more protected nodes fits with the organization of emergency and rescue operations. Whereas the rescue teams typically move on foot, the on-site rescue management is normally vehicle mounted nearby. The rescue management represents a common point of trust. These nodes may be better protected and less resource constrained than the ordinary nodes. However, the comments regarding the Z-H focus on CA availability and applicability for SAD and MAD operations applies to MOCA as well. The MP maintaining its own routing tables in parallel with a “standard” ad hoc routing protocol is superfluous and a waste of bandwidth.

---Secure and Efficient Key Management (SEKM)---In essence, SEKM [31] suggest the servers of MOCA form a multicast group. The aim is efficient updating of secret shares and certificates. A node broadcasts a certificate request to the CA server group. The server that first receives the request, generates a partial signature, and forwards the request to an additional $k+\alpha$ servers (not a true multicast). Only $k$ partial signatures are required. The additional ones are for redundancy in case some are lost or corrupted. SEKM does not state
how a server can tell it is the first to receive the refresh request and start the $k+a$ forwarding. On the whole, SEKM has the same features as MOCA. The required number of servers still has to be contacted, and the partial signatures returned.

**Ubiquitous Security Support (UBIQ)** - UBIQ [32] is a fully distributed threshold CA scheme. Similar to the partially distributed CA schemes Z-H, MOCA, and SEKM, it relies on a threshold signature system with a $(k,n)$ secret sharing of the private CA key. Differently from the partially distributed CA schemes, all nodes get a share of the private CA key. A coalition of $k$ one-hop neighbors forms the local CA functionality. It does not require any underlying routing protocol – only a node density of $k$ or more one-hop neighbors. Mobility may help finding the required number of CA nodes. UBIQ prescribes share refreshing.

The nodes earn trust in the entire network when they receive a valid certificate. Any node holding a certificate can obtain a share of the private CA key. A new secret share is calculated by adding partial shares received from a coalition of $k$ neighbors. The first nodes receive their certificates from a dealer before joining the network. After $k$ nodes have been initialized, the dealer is removed. The authors suggest that as the certification service is delivered within one-hop neighborhoods, some reliable out-of-bound physical proofs, such as human perception, can be used to authenticate new nodes.

Limiting CA service requests to one-hop neighborhoods is bandwidth efficient and good for the scalability. From a network point of view, the distributed trust management fits with both SAD and MAD operations scenarios. A local coalition can decide to let in nodes from different domains. A drawback in the rescue operations scenario is the possible requirement of human involvement. In addition, $k$ should be chosen carefully. A low value reduces intrusion tolerance. A large $k$ necessitates many neighbors. Reference [33] suggests more shares per node to succeed also with less than $k$ neighbors. In effect, this solution gives little else than reducing the value of $k$. Distributing the CA functionality boosts the availability of private key shares. Anyone capable of collecting $k$ shares or more can reconstruct the private CA key. Like any public key scheme relying on a trusted entity, there is no easy way to change the private/public CA key pair during operation.

In [34] it is argued that UBIQ may succumb to a Sybil attack [35] where a single node takes on more identities. With offline authentication of new nodes and the certificates
serving as proof of trustworthiness, this is hardly a realistic threat – at least not in settings like emergency and rescue operations. Secure and efficient revocation is an unsolved challenge.

--Autonomous Key Management (AKM) --AKM [27] provides a self-organizing and fully distributed threshold CA. With few nodes in the network, the scheme is parallel to UBIQ. Each node receives a share of the private CA key. As the number of nodes increases, a hierarchy of key shares is introduced. New nodes then receive a share of a share of the private CA key.

The root CA private/public key pair is bootstrapped by a group of neighbor nodes through distributed verifiable secret sharing [36]: Each of the n neighbors, chooses a secret value $S_i$, and distributes secret shares of this to the other neighbors using a $(k,n)$ secret sharing scheme\(^2\). Authentication is added offline. The sum of the individual secret values $S=(S_1+S_2+S_3+...+S_n)$ represents the private CA key. The corresponding public CA key equals $g^S$ (operations are mod prime $p$). Assuming the nodes publish the individual public values, $g^{S_i}$, the public key can be derived without revealing the private CA key by multiplying individual values $g^S=g^{S_1}*g^{S_2}...*g^{S_n}$. Figure 3 shows the principles.

The nodes (N1-N6) and their shares, $f(Ni)$, can be regarded as the leaves of a tree-structure. “R” in Figure 3 is a virtual node representing the private CA key. The probability of a compromise increases with more nodes holding a share of the private CA key. Therefore, when the number of share-holders reaches a certain level, the nodes split into smaller regional groups that set up a new regional key. Before splitting, the nodes N1-N6 hold shares $f(N1)$ - $f(N6)$ of the private CA key. Assuming the nodes N1-N3 decide to form a new group and N4 -N6 another; N1 distributes a share of its secret share $f(N1)$ to the other nodes in the new group. The others do the same with their key shares. The new regional secret of N1, N2, and N3 equals the sum of their shares $S'= f(N1)+ f(N2)+ f(N3)$, represented as virtual node “G” in Fig. 3.

\(^2\) This approach is contributory in nature. However, derivation of the individual private/public key pairs of the nodes is not. AKM is therefore classified as a distributive scheme.
When the number of shareholders in any region reaches the specified level, the region is split. Regions are also merged. With less than $k$ nodes, there are too few nodes to provide CA service. Certificates signed with regional keys have less assurance than those signed with the CA key. A high-assurance certificate requires partial signatures from nodes in different regions. The scheme assumes the network evolves from the nodes that initiated the AKM service.

MAD operations require nodes from one domain to be included in the other as key-share trees rooted in different private CA keys cannot be merged.

In AKM, each node maintains a CRL. AKM does not specify network-wide dissemination of revocation information. A certificate is revoked when at least $k$ neighbors have posted accusations against it. From a security point of view, it is questionable to what extent a certificate signed by the private CA key should be revoked by a group holding only a share of the private CA key.

AKM increases intrusion tolerance at the price of communicational cost. Nodes are assumed to disassociate with the previous region and associate with the new when they move from one region of the network to another. Implicitly, the nodes must maintain a view of the key hierarchy and be able to detect regional boundaries. With mobile nodes and unstable links, it is not evident how this can be implemented. The scheme requires the
nodes to collaborate on changes in regions and key hierarchy. Byzantine or faulty nodes may delay these operations. In scenarios like emergency and rescue operations, where the CA services primarily are needed for issuance of initial certificates and revocation, a hierarchical AKM with several regions represents a waste of bandwidth. For robustness and scalability, a single region is preferable. The scheme then equals UBIQ.

--Self-organized Key Management (PGP-A) -- In [34], Capkun, Buttyán, and Hubaux propose a fully self-organizing key management scheme (PGP-A) – a PGP [37] scheme adapted to ad hoc networks. The CA functionality is completely distributed. All nodes have equal roles. They generate their own private/public key pair and issue certificates to the nodes they trust. Certificates are stored in the nodes rather than in centralized repositories. PGP-A assumes trust is transitive (i.e., if Alice trusts Bob, and Bob trusts Charlie, then Alice should also trust Charlie). The nodes merge their certificate repositories, and try to find a verifiable chain of certificates. The Maximum Degree algorithm is suggested to construct a certificate graph with high connectivity even if the sizes of the users’ certificate repositories are small – due to the Small World phenomenon (the hypothesis that everyone in the world can be reached through a short chain of social acquaintances). Certificates are revoked through revocation messages from their issuer, or implicitly revoked at expiry time. Renewals require contact with the issuer. Certificates are also exchanged periodically between neighbor nodes. Evaluation of expiration times and periodical exchanges requires some sort of synchronization between the nodes. It is not evident from the paper how this synchronization should be established.

The periodic certificate exchanges and contact with issuers to have certificates updated is bandwidth consuming and scales badly. PGP-A implicitly requires an already running routing protocol. Trust could be established ad hoc through physical contact and key-exchange via a side channel. This enables improvisations suitable for both SAD and MAD operations. However, human interaction to keep network service running is undesirable.

Byzantine behavior or faulty nodes have limited power to prevent others from exchanging certificates. A compromised node only discloses the keys held by this node. Still, a compromised node could be used to issue certificates allowing other illegitimate nodes to gain access to the network.

There is only a probabilistic guarantee that a chain of trust can be found between parties
wishing to communicate. On the other hand, trust transitivity combined with the reliance on
the small world phenomenon implies that everyone will soon end up trusting everyone. The
result is no intrusion resistance. An alternative would be to restrict the maximum number of
hops, and allow the nodes to differentiate the level of trust they put in the various
certificates, as suggested in COMP [38].

---Composite key management for ad hoc networks (COMP)---

COMP [38] combines MOCA’s [29,30] partially distributed threshold CA with PGP-A [34] certificate-chaining.
The aim is higher security than obtainable with PGP-A, and increased availability of the
CA service compared to MOCA. Nodes that have been certified by the CA are allowed to
issue certificates to others. Nodes requesting a certification service should first try the
MOCA CAs. If this fails, they should search for neighbors that have been certified by the
CA. Depending on configuration, nodes with longer certificate chains to the CA may also
be entitled to issue certificates to others.

Each certificate in COMP includes a confidence value reflecting the level of confidence
the certificate issuer has in the binding between node identity and key (0=no trust, 1=full
trust). Multiplication of the confidence values gives a measure for the level of trust in a
certificate chain. Short certificate chains are generally preferred over long ones. The
probability of one or more compromised nodes in the chain grows as the length of the chain
increases. Similarly to PGP-A, COMP assumes a level of trust transitivity. However,
signing a certificate, verifying you believe a key belongs to a certain identity, does not
necessarily have to mean you also trust this identity to correctly sign certificates of others.

The confidence values enable fine grained evaluation of trust, and the nodes do not have
to trust the CA fully. However, deciding a proper confidence level is difficult. COMP does
not state how the certificate issuers should accomplish this. Byzantine or compromised
nodes may in any case assign full trust to untrustworthy nodes. Nevertheless, intrusion
tolerance is increased compared to pure PGP-A as COMP restricts the maximum length of
the certificate chains.

Offline authentication typically includes human interaction, which is cumbersome in the
setting of emergency and rescue operations. Interaction with one neighbor is less
demanding than the UBIQ requirement for involvement of several neighbors, though. Still,
COMP scales no better than MOCA as nodes requesting CA service should first try the
MOCA CAs. Transfers of certificate chains limit the scalability additionally.

In MANETs for applications like emergency and rescue operations, the CA will be expected primarily to be needed to issue and revoke certificates. Periodical updates of the certificates should not take place online during a rescue operation. Revocation is not addressed by COMP. It is reasonable that the node that issued a certificate is entitled to revoke it. But empowering single ordinary nodes to revoke certificates issued by the CA solely because they hold a certificate signed by the CA, renders the system vulnerable to compromised and Byzantine behaving nodes. Allowing a single node to issue certificates contradicts the purpose of the distributed CA.

A search for neighbors certified by the CA in order to obtain an initial certificate requires knowledge of the public CA key. Hence, at some point there should have been an authenticated channel between the searching node and the CA. The initial authenticated channel is typically obtained through physical contact, or a short-range side channel. A natural question for the node asked to provide CA service is then: Why did the requesting node not receive its certificate through the authenticated channel simultaneously?

In SAD operations the certificates could be pre-distributed. In MAD operations it may be hard for a node to verify whether a certificate from the other domain really has been certified by the correct CA or not.

**--Mobility-based key management scheme (MOB)** -MOB [39,40] seeks to mimic human behavior: if people want to communicate securely, they just get close to each other in order to exchange information. Security associations are established between pairs of nodes that get close. The scheme can be fully self-organizing (MOB-so) or rely on an offline authority (MOB-a). MOB-so can be based on symmetric or public keys. MOB-a is intrinsically public key based.

A major difference between MOB-so and MOB-a lies in the level of human involvement. In MOB-so, the users should authenticate the communicating peer physically before they establish a security association. The security credentials, triplets, are then exchanged over a secure (short-range) side channel. The triplets include user identifier, key and node address. The nodes also sign and exchange a statement that proves a security association has been established between the two. MOB-so accepts one level of transitivity in trust: security associations can be established through friends (i.e., nodes that have security associations to
both nodes in question). MOB-a assumes pre-distributed certificates, and suggests the exchange of security credentials is restricted to one-hop neighborhood.

In both MOB-so and MOB-a, only the keys held by the specific node are compromised when a node is captured. Byzantine behavior or faulty nodes do not inhibit others from exchanging security credentials. The off-line authority assumed by MOB-a implies no revocation. The authors suggest compromised nodes should revoke their own certificates. However, it can be hard to tell whether a compromise has taken place or not. Revocation on suspicion represents a vulnerability. It may be a threat to availability. Furthermore, if the node has been captured, it may no longer operate according to protocol. With MOB-so, it is left for the user to decide which of its security associations are no longer valid and what friend nodes have turned into enemies.

The MOB schemes are bandwidth efficient in the sense that security credentials are only exchanged within one-hop neighborhoods. Still, the scalability is limited. The MOB schemes imply a long delay to establish security associations with all communication partners. This is also unsuitable for emergency and rescue operations.

The designers suggest MOB-a for routing security and lower layers, and MOB-so for the application layer. MOB-a fits the SAD operations setting. MAD operations would require an on-line certificate authority to distribute cross-certificates of the merging domains.

MOB-a brings little achievement over pre-distributed certificates without restrictions on certificate exchanges. Depending on routing protocol, confining certificate exchanges to one-hop neighborhood may inhibit efficient network formation. There is no security achievement from such a restriction. The signature of the authority ensures the validity of the certificate no matter from whom the certificate was received. The assumption of MOB that no one should communicate securely with parties they have not been close to, contradicts the evolution of PKI.

--Identity-based public key (IBC-K) --Identity-based Cryptography [5], introduced by Shamir, removes the need for certificates. Identities are typically short – at least compared to certificates with a size of several kilobytes. Assuming information that is by default transferred in the routing messages can be used as the public key, identity-based schemes may scale better than the traditional certificate-based approaches. This makes Identity-based protocols interesting for bandwidth-limited ad hoc networks.
Shamir constructed an identity-based signature (IBS) scheme. To verify a signature, it is enough to know the ID of the sender plus the public system parameters. The public system parameters are defined by the private key generator (PKG) during system setup. The public system parameters include the public key of the PKG and information about the message space. The PKG also generates the private signature keys corresponding to the user IDs. Figure 4 shows a sketch of Shamir’s IBS scheme. During the setup phase, the PKG, chooses a secret master key and generates the corresponding public system parameters. Afterwards, in the extraction phase, it issues private keys. The private keys are uniquely given by the IDs and the PKG private master key.

Several IBS schemes have later been proposed. Some examples are found in [41-43]. Boneh and Franklin [44] introduced the first practical identity-based encryption scheme (IBE). This scheme has later been extended by Lynn [45] to provide message authentication at no additional cost. The ciphertext itself serves as the message authentication code.

Integration of identity-based signature and encryption schemes (IBSE) is studied in [46]. The latest progresses in IBE encompass strengthened security. Boneh and Boyen [47] suggested the first IBE scheme proven to be secure also in security models without random
oracles. Waters suggests a more efficient version in [43]. However, the IBE, IBSE and IBSC schemes presuppose pairwise communication. None are applicable for network layer one-to-many signing and verification of routing information.

The PKG represents a single point of failure. If the private master key of the PKG is compromised, the entire system is compromised. To counter this, Boneh and Franklin [44], suggest spreading the PKG master-key over more locations using threshold cryptography.

Khalili, Katz and Arbaugh [48] propose a key-management technique (IBC-K) for ad hoc networks combining identity-based cryptography with threshold cryptography [25]. The nodes that initialize the ad hoc network form a threshold PKG, spreading the PKG private master key over the initial set of nodes by a \((k,n)\) threshold scheme. This eliminates the PKG as a single point of failure, and adds intrusion tolerance. It makes the service robust in the sense that an adversary must compromise minimum \(k\) nodes in order to recover the secret master key. It also reduces vulnerability as the service is available as long as \(k\) correctly behaving PKG nodes are within reach.

In order to receive the private key corresponding to some identity, a node must present its identity to \(k\) (or more) of the \(n\) PKG nodes. The node receives a share of the private key from each of them. With \(k\) correct shares, the node can then compute its personal private key.

In SAD operations the private keys and system parameters could be handed out from an off-line PKG in the preparation phase. The IBC-K approach fits both SAD and MAD operations scenarios as it makes self-configuration of the PKG service possible. However, offline and mutual authentication is in any case required between entering nodes and any PKG node issuing a private key or key share. A secure channel is required. This implies physical contact or a short-range dedicated communication channel. Multihop connectivity is not good enough. It would enable passive eavesdropping as well as man-in-the-middle attacks.

When time is scarce, physical interaction with a number of geographically distributed PKG nodes is no good solution. Hence, for scenarios like emergency and rescue operations, a single PKG (e.g., located at the on-site rescue management) or a hierarchy of PKGs [49] would be more acceptable.

Explicit key revocation remains an unsolved problem. There is no easy way of
distributing revocation lists (withdraw IDs) and make sure all nodes are informed. Another
alternative is to change the PKG master key and system parameters. All private keys are
derived from these parameters. In essence, an update of the PKG key makes all keys in the
system obsolete.

**Summary of Public Key Schemes** – The capabilities of the public key schemes are
summarized in Table 2. IBC-K, making certificate exchanges superfluous, is an interesting
candidate for ad hoc networks. The reliance of a PKG makes it best suited for SAD
operations. Depending on whether the security policy demands centralized trust
management or not, IBC-K or COMP/UBIQ fits better in case of MAD operations.

2) Symmetric Schemes

---Pre-shared group key (PSGK): This is the old and well-proven key-management
scheme with a key distribution centre predistributing a symmetric key to all members of the
group. A key distribution centre could also provide pairwise unique keys, but the focus
here is on group keys. The symmetric group key can be used to “sign” routing information
with a cryptographic checksum – MAC (Message Authentication Code).

PSGK lacks intrusion tolerance in the sense that security succumbs to a single captured
node. But if the security policy allows it, it is a simple solution. Assuming an offline key
distribution center and predistributed keys, the scheme scales well. It is immune to faulty
nodes and Byzantine behavior. MAD operations would require a means of transferring the
group key from one node to another (via a location limited optical channel or similar).
Authentication should be added off-line. With a single group key, there is no easy way to
exclude compromised nodes.

PSGK was not designed specially for ad hoc networks. It is included here as several of
the symmetric schemes studied basically represent extensions to this scheme.

---SKiMPy [50] –SKiMPy is designed for MANETs in emergency and rescue operations.
It seeks to establish a MANET wide symmetric key for protection of network-layer routing
information or application layer user data. On MANET initialization, all nodes generate a
random symmetric key and advertise it within one-hop neighborhoods through HELLO
messages. The *best* key (i.e., the one with lowest ID number, freshest timestamp etc.) is
chosen as the local group key. The best key is transferred to the nodes with worse keys
through a secure channel established with the aid of pre-distributed certificates. The procedure is repeated until the “best” key has been shared with all nodes in the MANET. Once established, the group key serves as proof of trustworthiness. SKiMPy proposes periodical updates of the group key to counter cryptoanalysis. The updated keys are derived from the initial group key.

SKiMPy is bandwidth efficient in the sense that nodes agree on the best key locally. There is no need for an already running routing protocol as the key information is exchanged between neighbors only. SKiMPy implies a delay to spread the best key to all nodes. Still, the currently best local key can be used to communicate securely until the “ultimate” key is received.

Byzantine behavior or faulty nodes may disturb local key agreement, for example, by announcing a better key but not responding.

SKiMPy is designed for SAD operations. MAD operations would require some means of spreading cross-certificates. The authors indicate persons with special roles or ranks could be empowered to administer certificates. However, online revocation is not possible before the network has been initialized. As the network is initialized, the symmetric group key is also established. Once the symmetric key has been received, there is no efficient way to expel the node from further participation. The group key (or a key derived from it) now serves as proof of trustworthiness. Thus, SKiMPy adds complexity compared to PSGK, but does not increase the security accordingly.

---

**Self-Healing session key distribution (S-HEAL)** [51] –S-HEAL is a symmetric group key distribution scheme with revocation, designed for networks with unreliable links. The concept demands preshared secrets and a group manager that broadcasts the current group key $K$ “masked” with a polynomial $h(x)$; $f(x)=h(x)+K$. Individual secrets $h(i)$ are pre-distributed ($i$ refers to node ID). Each member node can then extract the current key by evaluating the received expression at $x=i$, and subtracting the secret value; $f(i)-h(i)=K$. All operations take place in a finite field $F_q$ where $q$ is a prime larger than the number of nodes.

Revocation is enabled by replacing the polynomial $h(x)$ with a bivariate polynomial $s(x,y)$. The group manager now broadcast the current key $K$ masked as $f(N,x)=s(N,x)+K$. In order to extract the key, the nodes must first recover the polynomial $s(x,i)$ and evaluate it at $s(N,i)$. Then they must subtract the result from the received $s(N,x)+K$, evaluated at $x=i$; $K=$
\( f(N,i) - s(N,i) \).

The thought is that only non-revoked nodes shall be able to recover the polynomial \( s(x,i) \). Given \( s \) of degree \( t \), \( t+1 \) values are required to find \( s(x,i) \). The value \( N \) and the individual secrets, \( s(i,i) \) are pre-distributed. The other \( t \) values, \( s(r_1,x), s(r_2,x) ... s(r_t,x) \), that are required to reveal \( s(x,i) \), are incorporated in the key update message from the group manager. If the revoked nodes are included in the set \( \{ r_1, r_2...r_t \} \), these nodes will only acquire \( t \) of the required \( t+1 \) values. Consequently, they will not be able to extract the new group key. The scheme enables revocation of maximum \( t \) nodes.

A main feature of S-HEAL is its self-healing property. Nodes that lose one or more key distributions can still reveal the missed keys. Each key update message includes shares of all of earlier as well as all possible future keys. The key shares received before are complementary to the shares received after the key has been distributed. Assuming \( p(x) \) is the share received before \( K \) is distributed, the share received in key update messages after \( K \) has been distributed equals \( K-p(x) \). Hence, missed keys can be derived by combining shares received before the lost update with shares received after the lost update. Whereas the self-healing feature may be of great value in mail systems and similar applications, network layer routing information has only instant value. Hence, retrieving earlier keys is of little interest. Further details are therefore left out.

S-HEAL’s reliance on a group manager – possibly multi-hops away – to provide the initial group key, makes it inapplicable for protection of routing information. The group key is needed in order to bootstrap the network service, but S-HEAL demands an already running network service to distribute the group key. Nevertheless, S-HEAL could potentially be used for revocation and re-keying, assuming a protected network service has been bootstrapped with an initial pre-distributed group key (PSGK). It would improve intrusion tolerance compared to pure PSGK. Robustness to packet losses could be increased by periodically retransmitting the latest key update rather than waiting for the next key update as implied by [51].

Regarding scalability, the message sizes and number of key update messages are independent of the number of nodes in the network. The size of the key updates is only proportional to the size of the polynomials (if self-healing is left out.)

Predistributed individual shares are acceptable for SAD operations, but incompatible with
merging of nodes from different security domains. Thus, S-HEAL has limited applicability in MAD operations.

Missing source authentication of the broadcasts from the group manager is a shortcoming. A MAC generated by the previous group key could easily be added. Still, a Byzantine behavior node could potentially transmit garble, claiming to be the next key from the group manager, and cause disruption.

--Logical Key Hierarchy (LKH) --Group keys can be updated brute force: A group manager distributes the new group key, encrypted with a separate (individual) key for each node. In essence, LKH represents a family of schemes that improve the scalability of this brute force method by organizing the keys into a logical hierarchy and giving the nodes additional keys.

LKH was introduced by Wong, Gouda, and Lam [52] and Wallner, Harder, and Agee [53]. The concept is illustrated in Fig.5. All group members (N1-N8) possess the group key $K_{12345678}$. The subgroup key $K_{1234}$ is shared by members N1- N4, and $K_{12}$ is common to N1 and N2. $K_1 – K_8$ refer to the individual keys. Assuming node N8 is to be revoked; all group and subgroup keys known to N8 ($K_{12345678}, K_{5678}$ and $K_{78}$) should be updated. N7 shares all intermediate keys from the leaf to the root with N8. N7 must therefore receive the updated keys encrypted with its individual key. The new group key and subgroup key can be distributed to N5 and N6 encrypted with their key in common; $K_{56}$. To N1-N4, the group manager sends the new group key $K_{1234567}$ encrypted with $K_{1234}$. Thus, bandwidth and computational cost is saved compared to updates encrypted with the individual keys.

The key tree can be binary or k-ary, and balanced or unbalanced.

---

Figure 5. Logical key hierarchy
Whereas the basic LKH scheme was not designed specifically for ad hoc networks, Rhee, Park and Tsudik [54, 55] suggest a LKH scheme for hierarchical ad hoc networks. They propose that the group manager functionality is distributed over several managers, each controlling different *cells* of the network. The approach is cellular and infrastructure based rather than ad hoc. The nodes are fully dependent upon the cell managers. Each cell has a different group key.

The nodes must contact the cell manager to receive the key when they move from one cell to another. In other words, the nodes must be able to detect cell boundaries and be within communication range of a cell manager. In addition, the scheme requires that cell managers communicate during “key hand-off” from one cell to another. The intention of the scheme is to limit re-keying to part of the network. The price is reduced robustness and increased bandwidth consumption. The scheme is inapplicable for MANET use.

A number of other refinements of the basic LKH scheme [52, 53] focusing on communicational and computational cost, have been proposed. OFT [56,57], OFC [58], ELK [59], LKH+ [56], EBHT [60], LKH++ [61], Poovendran and Baras [62], and the Internet-draft by Selcuk, McCubbin and Sidhu [63] all propose different ways to reduce communication overhead –primarily focusing on message sizes. Of these schemes, only LKH++ is designed for wireless networks.

ELK reduces the size of the key update messages by sending only part of the key plus a key verification value. The receivers must search brute force for the remaining part of the key. The verification value is used to decide whether the correct key has been found or not. LKH+ and EBHT suggest new keys are derived by applying a one-way function to the old key(s) when members are added. In OFT, OFC, and LKH++ the keys of parent nodes are related to keys of their child-nodes through one-way functions. After a group change, the group manager only sends enough information to enable the nodes to compute the rest of the updated keys themselves. Poovendran and Baras [62] show that the overhead can be reduced by placing nodes that are most likely to be revoked as close as possible to the root, that is, giving them only a minimum number of keys. Similar ideas are also studied in [63]. In a practical situation it can be hard to decide which nodes are most likely to be compromised. Furthermore, in ad hoc networks, the number of messages may be more
devastating than the size of the messages [64]. Hence, the actual gain of the proposed refinements is not evident. Simulations are required in order to judge which approach is the better.

LKH++ claims to reduce the number of messages as some of the nodes will be able to calculate the new key by themselves. The keys of parent nodes are related to keys of their child nodes through one-way functions. According to LKH++ and referring to Fig. 5, $K_{12}$ equals a hash of $K_1$, and $K_{1234}$ represents a hash of $K_{12}$. Consequently, children to the left will be able to calculate the new key of their parent node. The others receive it from the group manager. Still, the nodes to the left must also be made aware that a key update is required. LKH++ does not address how to do this.

In an ad hoc network, rekeying every time a new node join or leave the network is unnecessary and undesirable. Routing information has only instant value. Backward and forward secrecy on joining and leaving nodes is of little importance. However, LKH may be of interest as an extension to PSGK for revocation purposes. Assuming the network service has been initialized (with the aid of the pre-distributed group key); LKH could be used to expel compromised nodes. A static tree, large enough to hold the keys of all anticipated members, would be required in order to avoid re-keying when new nodes are added. This may be possible for SAD operations. MAD operations would require the merge of two (or more) trees and necessitate re-keying.

For (infrequent) revocations in ad hoc networks for emergency and rescue operations, robustness is even more important than communicational and computational cost. In the basic LKH scheme, innocent nodes that miss the update from the group manager may be cut off. Periodical retransmissions of the last update(s) could help. In ELK, the group manager sends repeated hint messages that enable nodes that lost the key update to calculate the key. Forward error correction codes (FEC) on key updates, as suggested by Wong and Lam [65], enable correction of bit errors, but does not help nodes that missed the entire update.

The group manager represents a single point of failure. Replication of the group manager for reliability and performance, as suggested in [52], is of limited value for ad hoc networks. Replication demands synchronized servers and increase the number of targets for security attacks.
A general weakness of schemes that rely on symmetric keys only, is the missing possibility for source authentication. Byzantine nodes may pose as group manager and cause disruption. Reference [52] proposes authentication through digital signatures. The basic key management problem then reverts to public key distribution.

---Probabilistic Key Pre-distribution (PRE)---PRE [66] assumes WSN nodes outfitted with a preinstalled key ring, that is, a set of keys drawn randomly from a large pool of keys. When bootstrapping the network, the nodes broadcast the identifiers of the keys in their key ring. A wireless link is established between nodes only if they share a key. Hence, resilience to Byzantine behavior and faulty nodes is fine. The scheme relies on a controller node (base station) to broadcast a signed list of the key identifiers to be revoked.

A number of probabilistic key pre-distribution schemes for WSNs have been proposed. In [67], Chan, Perrig, and Song suggest extensions to [66] that increase the resilience against node capture. It requires \( q \) common keys \((q > 1)\) instead of just a single one to establish a connection. Liu and Ning [68] propose probabilistic pre-shared polynomials for establishment of pairwise keys in WSNs. Polynomial sharing increases resilience to captured nodes. A trusted entity defines a bivariate polynomial, \( f(x,y) \) with the property \( f(x,y) = f(y,x) \). Secret polynomial shares; \( f(i,y) \) are predistributed to each sensor node, \( i \). Any two nodes, \( i \) and \( j \), can set up a pairwise key by evaluating the polynomial at \( f(i,j) \) and \( f(j,i) \), respectively. Similarly, Du, Deng, Han and Varshney [69] suggest another scheme relying on probabilistic preshared polynomials for pairwise keys in sensor networks. In [70], Du et al. suggest use of deployment knowledge to increase the probability that two nodes find a common secret key. The latter may be possible in a WSN with planned positioning of sensors, but not in a MANET.

Zhu, Xhu, Setia and Jajodia [71] propose probabilistic key predistribution combined with secret sharing to set up pairwise exclusive keys in MANETs. A node, wishing to communicate securely with another, picks a secret symmetric key. It then sends shares of this secret symmetric key, encrypted with different pre-distributed keys to the opposite party (i.e., the shares are sent through different logical paths). Assuming the aggregated set of predistributed keys used are known to the two nodes in question only, no other nodes will be able to decrypt enough shares to reveal the secret symmetric key. Depending on configuration, the scheme may produce a large number of messages. Reference [71] claims
it is desirable to trade computation for communication in ad hoc networks. This assumption
does generally not hold.

The idea of the key ring of PRE is intrusion tolerance. The price is availability. There is
only a probabilistic assurance that a node actually will share a key with one or more
neighbors and be able to bootstrap communication. Emergency and rescue operations,
where availability is a number one concern, would require a key ring large enough to
achieve close to zero probability of failure. The consequence is intrusion resistance reduced
to a level comparable to a predistributed group key (PSGK). This contradicts the intension
of the scheme. The applicability and scalability for network layer security is limited.
Different keys in common with the various neighbors imply more signatures for each
routing message. End-to-end signatures on routing messages to be flooded are precluded.

-- Security Protocols for Sensor Networks (SPINS) [72] --The SPINS security protocols
for WSNs assume preinstalled individual (pairwise) keys between the sensor nodes and a
base station. Nodes that want to communicate securely, request the base station for a
common key. The base station returns the key, encrypted with their individual keys. This
scheme demands an already running routing protocol and reliable access to the base station.
It is inapplicable for the purpose of protecting routing information in a traditional MANET.

SPINS also includes a scheme for authenticated broadcast; µTESLA, and describes how
this can be used to provide an authenticated routing protocol for sensor networks. µTESLA
relies on a predistributed commitment (i.e., the last key of a one-way key chain) and
delayed disclosure of subsequent keys in the key chain. The key chain can be derived by
repeated hashes of an initial random key. The key used at time $i$ equals a hash (or similar
one-way function) of the key used at time $i+1$. The commitment enables the nodes to
verify that later disclosed keys originate from the claimed source; repeated hashes of the
disclosed key should return the commitment. To send an authenticated packet, the sender
computes a message authentication code (MAC) with a key that is secret at that point in
time. The receiver stores the message until the key later is disclosed. The nodes must be
loosely time synchronized and know the key disclosure schedule. Otherwise, adversaries
could forge packets as the receiver would not know whether the key used to calculate the
MAC of an incoming packet had been disclosed or not.

The SPINS authenticated routing protocol discovers routes from the nodes to the base
station with the aid of $\mu$TESLA key disclosure packets flooded from the base station. The sender, from which a node first received the valid $\mu$TESLA packet, is set as parent node in the route to the base station. The predistributed commitment enables the nodes to verify that the received packet originated from the base station.

This may work for communication from sensor nodes to a base station. The same technique cannot be used in a traditional MANET with a scattered communication pattern. One possibility would be to preload all nodes with commitments of the key chains of all other nodes. This would allow any node to authenticate the messages from any other node. Intrusion tolerance would be fine, and robustness to Byzantine behavior and fault nodes is good. However, the nodes would have to be loosely time synchronized and know the key disclosure schedule of all other nodes. The solution would give little flexibility and scale badly. MAD operations and late registered nodes in SAD operations would be precluded. Furthermore, delayed key disclosure is problematic in the setting of mobile nodes and rapidly changing network topology. Altogether, the SPINS key management scheme and authenticated routing protocol is inapplicable for protection of routing information in traditional MANETs.

--GKMPAN [73] --GKMPAN is designed for secure multicast in ad hoc networks. It is basically a revocation and re-keying scheme for PSGK, founded on PRE [66] and $\mu$TESLA [72].\(^3\) GKMPAN assumes a pre-distributed group key plus a pre-distributed commitment. The group key is used to protect multicast communication. The commitment is used for authentication of revocation messages from the key server. In addition, GKMPAN assumes each node is equipped with a preinstalled subset of symmetric keys drawn from a large key pool.

In contrast to PRE, the keys in the key set are determined from the ID of the node. On revocation, the key server issues a revocation message containing the ID of the revoked node. All keys in the key set of the revoked node should be erased or updated. Any node can automatically tell from the ID which keys to revoke. The revocation message also

\(^3\) Actually, [73] claims TESLA [74] is used for authentication. At the same time [73] assumes a pre-distributed commitment, and states that only symmetric key techniques are used – which implies $\mu$TESLA. The difference is subtle. $\mu$TESLA relies on a pre-distributed commitment. TESLA assumes the initial packet is authenticated with a digital signature.
identifies a key that is not in the key set of the revoked node, to be used as “update key.”

A new group key is derived from the old one with the aid of a keyed one-way function; the old group key is used as data input and the “update key” as key input. The output is the new group key. Nodes that have the “update key” in their key set, can calculate the new group key without assistance. The others receive it from their parent node, encrypted with one of the (non-revoked) keys in their key set. It is distributed through a multicast tree rooted at the key server. The validity of the revocation message cannot be checked before the key server later discloses the key that was used to compute the message authentication code.

In order to avoid potential disruptions, the old and new group key should coexist until all nodes have received the new group key. However, there is no easy way to make sure all nodes have received the new group key. Byzantine behavior or faulty nodes may inhibit efficient exclusion and re-keying. The reliance on a key server and time synchronization are other vulnerabilities. GKMPAN scales fairly in the sense that new group keys can be calculated by the nodes themselves or transferred locally. In addition, GKMPAN increases intrusion tolerance compared to PSGK as it enables node exclusion. The price is reduced availability. Innocent nodes may be expelled if all their keys happen to be in revoked key sets. This is not acceptable in settings like emergency and rescue operations.

--Secure Pebblenets (PEBL) [75]: Pebblenets refer to large ad hoc networks where the nodes are called pebbles due to their small size and large number (e.g., WSNs). The aim of PEBL is protection of application data. It establishes and updates a network-wide traffic encryption key, TEK. At the network layer a preinstalled group key guarantees the authenticity of a pebble as a member of a group. Hence, PEBL can be regarded as an extension to PSGK. The assumption is that only nodes possessing the group key are capable of encrypting and decrypting HELLO messages correctly. Furthermore, PEBL assumes the pebbles organize into clusters of one-hop neighbors. Each cluster selects a clusterhead node. The clusterheads establish a backbone, and compete to become the key manager. The key manager generates the traffic key, TEK, which is intended for encryption of application-layer data traffic. The TEK is distributed from the key manager to the regular nodes through the clusterheads. It is updated periodically. Each TEK update is preceded by a reclustering and new selection of clusterheads. This rotation of the clusterhead role is to
avoid exhaustion of the nodes acting as clusterheads, and to account for mobility. Nodes that were one-hop neighbors when the cluster was formed may have moved out of the neighborhood.

Nodes that do not behave according to the protocol may disturb cluster formation and TEK updates. PEBL offers no protection against replay or intrusion. PEBL security succumbs to tampering. Both network layer HELLO messages and TEKs are all protected by keys derived from the group key. Anyone possessing the group key will be able to participate in the TEK updates. PEBL in its entirety, with cluster formation and periodic TEK updates, is bandwidth consuming, demands synchronization and makes availability assumptions that renders it not suited for MANET use.

---

**Key Infection (INF)** [76] –INF is intended for WSNs. The scheme assumes static sensor nodes and mass deployment. INF sets up symmetric keys between the nodes and their one-hop neighbors. The security is based on surprise: It relies on the assumption that during the network deployment phase, any attacker is only able to monitor a fixed percentage of the communication channels. At bootstrap time, every node simply generates a symmetric key and sends it in the clear to its neighbors. A *key whispering* approach is used, i.e. the key is initially transmitted at a low power level. The transmission power is then increased until the key is heard by at least one of its one-hop neighbors and a reply is received. INF is simple, self-organizing, and robust to Byzantine behavior and faulty nodes. It is bandwidth efficient, and scales well. However, the security is weak. INF is vulnerable to eavesdropping during key whispering. In addition, there is no authentication of the communicating parties. INF’s “security through surprise” fails for MANETs where static nodes and instant mass deployment is no option.

---

**Localized Encryption and Authentication Protocol (LEAP)** [77] is designed for static WSNs. LEAP suggests different keys for different purposes. It requires a number of pre-distributed keys. Predistributed *individual keys* are used for communication between sensor nodes and the base station. A preshared *group key* is applied for protection of broadcast information from the base station. A preinstalled network wide *initial key, K*, is used to derive *pairwise keys* for secure communication between one-hop neighbors.

During neighbor discovery immediately after deployment, each node $n$ derives its master key, $K_n$. The master key is derived as a function of its node ID and the initial key;
$K_n=f_k(ID_n)$. The master key is used to “sign” HELLO messages. Any node that knows the initial key is able to calculate the master key of any other node ID. Hence, each node can verify the HELLO messages received from its neighbors. The node then calculates the pairwise keys shared with its neighbors, $v$, as a function of their master key and the node ID; $K_{nv}=f_{kv}(ID_n)$.

Intrusion tolerance is obtained under the assumption of stationary nodes; the network key is erased after the pairwise keys have been established. Nodes that have erased the network key can no longer establish pairwise keys. New nodes can still be added though. As the new nodes have not yet erased the group key, they can set up pairwise keys with their neighbors. When a node is captured, only the keys held by the captured node are compromised.

The pairwise keys are used both to secure ordinary data and to distribute cluster keys. The cluster keys are employed for secure local broadcasts. Any node simply generates a cluster key and sends it to all neighbors, encrypted with the respective pairwise keys.

Whereas LEAP may work in a static sensor network, the heart of this key management scheme – the setup of pairwise keys – will not work in a traditional ad hoc network. Deletion of the initial key is incompatible with mobile nodes and constantly changing network topology. Evaluating the scalability of LEAP in MANETs makes little sense, as pairwise key set up is precluded after the initial key has been erased.

**Distributive symmetric Schemes-Summary** – Table 3 gives an overview of the capabilities of the distributive symmetric key management schemes. The WSN key management schemes generally assume static nodes, mass deployment, node-to-base station communication pattern or are designed to establish pairwise keys. Their aim and assumptions render them inapplicable for protection of routing information in traditional ad hoc networks with mobile nodes. PSGK, or PSGK extended with S-HEAL or LKH for revocation, appear to be the most promising alternatives of the symmetric schemes.

### IV Conclusions

We find it useful to classify key management schemes for MANETS (mobile ad hoc networks) as either *contributory* or *distributive*. Figure 6 provides an overview of the schemes surveyed in this paper. Distributive schemes based on symmetric techniques are
either intended for traditional MANETs or for wireless sensor networks (WSN). Distributive key management based on asymmetric cryptographic techniques may take on the standard distinction between certificate-based and identity-based schemes.

We were not able to select one single scheme that is intrinsically superior to the others in our comparative work. A general observation is that none of the proposed key-management schemes for MANETs are truly effective for all MANET scenarios. The application must be taken into consideration at current state of the art. There is a lack of reported attention to the challenges presented by the concrete limitations of communication capacity in MANETs. The performance evaluations tend to be restricted to computational complexity considerations, whereas the practical constraint of MANETs is more likely the communication capacity rather than the computational power and energy consumption. Both the size and the number of messages are important. The number of messages can have more impact than the size due to overhead introduced in lower communication layers.

The optimal combination of bandwidth efficiency and robustness against link loss under a given power consumption should be sought in future key management proposals. Also, secure and efficient key revocation remains an open challenge in MANETs.

![Figure 6. Key management schemes surveyed](image-url)
Acknowledgements

We thank the anonymous reviewers for all their suggestions and constructive criticism. We also thank Leif Nilsen, University of Oslo – University Graduate Centre/Thales Group Norway for valuable discussions and comments on our work. A special thanks to Prof. Loren Olson at the University of Tromsoe for proof-reading the manuscript.

References


<table>
<thead>
<tr>
<th>Characteristics</th>
<th>D-H</th>
<th>ING</th>
<th>B-D</th>
<th>H&amp;O</th>
<th>A-G</th>
<th>CLIQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two parties</td>
<td>Logical ring of nodes during D-H key agreement</td>
<td>Reduce the number of rounds to 3 by reliable multicasting</td>
<td>Reduce the number of rounds from $n$ to $d$ ($n=2^d$) by arranging nodes into hypercube</td>
<td>Password authenticated H&amp;O</td>
<td>Group changes through reliable multicast from group controller</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Applicability</th>
<th>Aim</th>
<th>Net</th>
<th>Authentification</th>
<th>Intrusion tolerance</th>
<th>Trust Management</th>
<th>Vulnerabilities</th>
<th>Availability assumptions</th>
<th>Byz. behavior &amp; Faulty nodes</th>
<th>Group changes</th>
<th>Scalability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>S.O.</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Peer</td>
<td>Y</td>
<td>NA</td>
<td>Poor</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>S.O.</td>
<td>PK</td>
<td>N</td>
<td>CA</td>
<td>MIM</td>
<td>Ring O</td>
<td>N</td>
<td>Re-key</td>
<td>Poor</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>Planned</td>
<td>S.O.</td>
<td>N</td>
<td>N</td>
<td>O, MIM, Byz. behavior</td>
<td>O + RM</td>
<td>O, Byz. behavior</td>
<td>N</td>
<td>Re-key</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>S.O.</td>
<td>S.O.</td>
<td>N</td>
<td>N</td>
<td>O, MIM, Byz. behavior</td>
<td>Hypercube O</td>
<td>O, Byz. behavior</td>
<td>N</td>
<td>Re-key</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>S.O.</td>
<td>S.O.</td>
<td>N</td>
<td>N</td>
<td>O, MIM, Byz. behavior</td>
<td>Hypercube O</td>
<td>O, Byz. behavior</td>
<td>N</td>
<td>Re-key</td>
</tr>
</tbody>
</table>

Byz.: Byzantine, G: Group Key (symmetric), GC: Group Controller, MIM: Man-in-the-middle, N: No/None, NA: not applicable, O: node ordering, P: Pairwise symmetric key, PK: public key, RM: reliable multicast, S.O.: self organizing, Y: yes
<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Z-H</th>
<th>MOCA</th>
<th>SEKM</th>
<th>UBIQ</th>
<th>AKM</th>
<th>PGP-A</th>
<th>COMP</th>
<th>MOB</th>
<th>IBC-K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group of servers</td>
<td>Group of servers forms</td>
<td>Most powerful nodes form</td>
<td>Threshold CA servers form multicast group</td>
<td>All nodes part of threshold CA</td>
<td>“Hierarchical UBIQ”</td>
<td>Anarchy: All nodes act as distinct CAs</td>
<td>MOCA+PGP-A</td>
<td>Move close for exchange of security credentials</td>
<td>No C, Key=ID</td>
</tr>
<tr>
<td>Aim</td>
<td>PD TCA</td>
<td>PD TCA</td>
<td>PD TCA</td>
<td>FD TCA</td>
<td>FD TCA</td>
<td>FD CA</td>
<td>PD CA</td>
<td>MOB-a: Offline CA</td>
<td>MOB-so: FD TCA</td>
</tr>
<tr>
<td>Intrusion tol.</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Fair</td>
<td>Fair/Good</td>
<td>Limited</td>
<td>Limited/Fair</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td>Trust Mgmt</td>
<td>CA</td>
<td>CA</td>
<td>CA</td>
<td>CA: k 1-hop neighbors</td>
<td>CA: k 1-hop neighbors</td>
<td>Nodes</td>
<td>CA + CA certified nodes</td>
<td>MOB-a: N (Off-line CA)</td>
<td>MOB-so: Nodes</td>
</tr>
<tr>
<td>Vulnerabilities</td>
<td>Combiner, CRL distr, CA key update</td>
<td>CRL distr, CA key update</td>
<td>CRL distr, CA key update</td>
<td>CRL distr, 1-hop neighbors &lt;k, (Sybil attack), CA key update</td>
<td>Regional changes, Revocation CRL distr 1-hop neighbors &lt;k, CA key update</td>
<td>Comprised nodes, CRL distr,</td>
<td>Comprised nodes, Distributed trust mmmnt, CRL distr, CA key update</td>
<td>Revocation, Delay due to restriction on Security credential exchanges, CA key update</td>
<td>IRL distr, PKG key update</td>
</tr>
<tr>
<td>Availability assumption</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Limited</td>
<td>Good</td>
<td>Limited</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td>Byz. behavior &amp; Faulty nodes</td>
<td>C + CRL</td>
<td>C + CRL</td>
<td>C + CRL</td>
<td>C + CRL</td>
<td>C+CRL/accusations+ Region size</td>
<td>C + CRL</td>
<td>C (+CRL)</td>
<td>C (+CRL)</td>
<td>IRL</td>
</tr>
<tr>
<td>Scalability</td>
<td>Poor</td>
<td>Limited</td>
<td>Limited</td>
<td>Fair</td>
<td>Limited</td>
<td>Poor</td>
<td>Limited</td>
<td>Limited</td>
<td>Fair</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>PSGK</th>
<th>SKiMPy</th>
<th>S-HEAL</th>
<th>LKH</th>
<th>PRE</th>
<th>SPINS</th>
<th>GKMPAN</th>
<th>PEBL</th>
<th>INF</th>
<th>LEAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-shared group key</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Establishment key on network formation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polynomial sharing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Key Hierarchy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probabilistic key distribution</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Table 3: Summary of symmetric Key Management Schemes |

<table>
<thead>
<tr>
<th>Applicability</th>
<th>Aim</th>
<th>GK</th>
<th>GK</th>
<th>Rev &amp; re-key**</th>
<th>Rev &amp; re-key**</th>
<th>Keys between subsets of 1-hop neighbors</th>
<th>PK and Authenticate d route to base station</th>
<th>GK, rev &amp; re-key</th>
<th>GK</th>
<th>Keys between 1-hop neighbors</th>
<th>GK, PK, cluster keys</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net</td>
<td>Authenti -cation</td>
<td>Off-line</td>
<td>C, KP</td>
<td>N</td>
<td>KP</td>
<td>KP</td>
<td>Off-line, µTESLA</td>
<td>KP</td>
<td>µTESLA</td>
<td>KP</td>
<td>N</td>
</tr>
<tr>
<td>Intrusion t tolerance</td>
<td>Poor</td>
<td>Limited</td>
<td>Fair</td>
<td>Fair</td>
<td>Fair</td>
<td>Fair</td>
<td>Limited</td>
<td>Poor</td>
<td>Poor</td>
<td>Limited</td>
<td></td>
</tr>
<tr>
<td>Security</td>
<td>Trust Mnmt</td>
<td>Off-line</td>
<td>Off-line / special nodes</td>
<td>G Mgr</td>
<td>G Mgr</td>
<td>Controller node</td>
<td>Base station</td>
<td>On-line Key Server</td>
<td>Off-line</td>
<td>N</td>
<td>Base station</td>
</tr>
<tr>
<td>Vulner-abilities</td>
<td>Tamper</td>
<td>Tamper, Rev, CA, Periodic key updates</td>
<td>G Mgr, colluding nodes&gt;1, Byz nodes</td>
<td>G Mgr, Byz nodes</td>
<td>Controller node</td>
<td>Base station, Synch</td>
<td>Key Server, synch, Byz nodes, rev of innocent</td>
<td>Tamper, Re-key, synch, Cluster head selection</td>
<td>Eavesdropping</td>
<td>Initial key, Node mobility, Base station</td>
<td></td>
</tr>
<tr>
<td>Avail-ability assumptions</td>
<td>N</td>
<td>N</td>
<td>G Mgr, Reliable key distribution</td>
<td>G Mgr, Reliable key distrib.</td>
<td>Key ring fits with neighbors’</td>
<td>Base station</td>
<td>Key Server</td>
<td>Synch, Full connectivit y during TEK establishment, Cluster head</td>
<td>1-hop neighbors&gt;1</td>
<td>No mobility</td>
<td></td>
</tr>
<tr>
<td>Byz behavior &amp; Faulty nodes</td>
<td>Good</td>
<td>Fair</td>
<td>Poor</td>
<td>Poor</td>
<td>Good</td>
<td>Good</td>
<td>Limited</td>
<td>Limited</td>
<td>Good</td>
<td>Limited</td>
<td></td>
</tr>
</tbody>
</table>

| Dynamic group changes | N | N | Re-key | Re-key | Key re-advertising | N | Re-key | Periodic TEK update | N | Re-key |

| Scalability | Re-source efficiency | Good | Fair | Fair | Fair | Limited | Poor | Fair | Limited | Good | *) |


*) assumes static nodes – scalability in MANETs with mobile nodes makes little sense
**) It is here assumed a predistributed group key and S-HEAL/LKH used for revocation
PAPER II: Analysis of IBS for MANET Security in Emergency and Rescue Operations

Anne Marie Hegland, Eli Winjum, Pål Spilling, Chunming Rong, and Øivind Kure

In Proceedings of the
IEEE International Workshop on Pervasive Computing and Ad Hoc Communications (PCAC ’06),
held in conjunction with the
IEEE International Conference on Advanced Information Networking and Applications (AINA ’06),
Vienna, Austria, April 18-20, 2006
Analysis of IBS for MANET Security in Emergency and Rescue Operations

ANNE MARIE HEGLAND* ELI WINJUM* PÅL SPILLING* CHUNMING RONG§ ØIVIND KURE†
*UniK, Kjeller, Norway, §UiS, Stavanger, Norway, †NTNU, Trondheim, Norway
{amh;ewi;okure;paal}@unik.no; chunming.rong@uis.no

Abstract

Protection of the network layer in mobile ad hoc networks (MANETs) imply cryptographically signed routing messages. Identity-based signature (IBS) schemes make bandwidth consuming certificate exchanges obsolete. User identifiers serve as public keys. However, long-term identifiers are required.

This paper analyzes the applicability of IBS schemes for securing routing information in MANETs for emergency and rescue operations. The Optimized Link State Routing protocol (OLSR) serves as example.

1 Introduction

With the wireless media of MANETs, there is no guarantee that malicious nodes do not mingle and interfere. A key issue is to judge whether a routing message comes from a trusted node or not. This calls for one-to-many signing and verification of routing messages. Traditional public key schemes require certificates that bind the public keys to identities. The newer identity-based signature (IBS) schemes make certificates obsolete. User identifiers serve as public keys. Assuming information that is by default transferred in the routing messages, e.g. IP addresses, can be used as public key; identity based public key schemes may scale better than traditional public key schemes. This makes identity-based protocols interesting for bandwidth limited ad hoc networks.

This paper analyzes the applicability of IBS for protection of routing information in MANETs for emergency and rescue operations. The aim is to pinpoint the consequences of an IBS system, rather than to track the best solution. OLSR [7] is
used as example. The rest of this paper is organized as follows. Section 2 describes the characteristics of emergency and rescue operations. Section 3 outlines identity-based signature schemes. OLSR is described in 4. IBS protection of ad hoc routing protocols is analyzed in 5. Related work is described in section 6. Conclusions and suggestions for future work are found in section 7.

2 Emergency and rescue operations

In emergency and rescue operations, an off-site Rescue Coordination Centre (RCC) establishes the on-site rescue management (OSRM) with an on-site coordinator (OSC) as head. The OSRM administers the on-site rescue team consisting of professional teams (typically including, police, health services and the fire brigade). Thus, the involved parties and roles are to a large extent predefined. Arbitrary civilian volunteers are not allowed to participate. This enables pre-configured security credentials. However, “ad hoc” partners from multiple security domains may be pulled in, e.g. in operations involving industrial fire brigades and cross-border operations.

Whereas the members of the professional units normally move on foot on the site of the accident, the OSRM is typically vehicle mounted nearby. The members of the on-site rescue team are equipped with wireless mobile nodes forming a MANET. There may be a communication channel back to fixed network through one or more MANET Gateways – possibly co-located with the OSRM. The MANET Gateways can be mobile base stations, TETRA base stations, or a satellite links, i.e. possibly low capacity links. It is assumed that all MANETs will be “stub” networks where the nodes only route traffic that originates and/or terminates in the particular MANET.

3 Identity-based signature schemes

Identity-based Cryptography [16] removes the need for certificates. User identifiers serve as public keys. The first IBS scheme, introduced by Shamir [16], is outlined in Fig.1. During the setup phase, the trusted entity, the Private Key Generator (PKG), chooses a secret master key and generates the corresponding public system parameters.
Fig. 1 Shamir’s IBS scheme

Afterwards, in the extraction phase, it issues private keys. The private keys are uniquely determined by the IDs and the PKG private master-key. Signature verification requires the ID of the sender plus the public system parameters. A merge of nodes from multiple security domains requires new public system parameters to be signed by OSRM/RCC (PKG) and flooded through the gateways.

Secure and efficient revocation is an unsolved challenge. Revocation decisions are naturally made by RCC or OSRM nodes with access to the gateways. Lists of revoked IDs, signed by the PKG, can thus be distributed through the gateways. If no gateways are accessible and operable, revocation is precluded.

4 The optimized Link State Routing Protocol (OLSR)

OLSR [7] is a proactive link state routing protocol designed for large and dense MANETs. Bandwidth consumption is optimized compared to classical flooding through the use of Multipoint Relay (MPR) nodes. Only nodes that are chosen as MPRs forward routing information. Each node selects MPRs from its 1-hop neighbors in such a way that all 2-hop neighbors are covered by at least one MPR node. Each node broadcasts its links and MPR selections through HELLO messages. HELLO messages are not forwarded. In contrast, Topology Change (TC) messages disseminate topology information throughout the network, and are generated and forwarded merely by MPR nodes. Routing tables are computed from the information exchanged through TC messages. An OLSR node may have several network interfaces, but there should always
be one main address identifying the node. The main address is used as the originator address in all OLSR messages [7]. Other OLSR interfaces are announced through Multiple Interface Declaration (MID) messages, and will be visible as source and destination addresses in the IP header. OLSR packets are encapsulated in UDP over IP. Interfaces to non-OLSR networks are declared through Host and Network Association (HNA) messages. Like TC messages, MID and HNA messages are flooded by the MPRs.

5 Protecting MANET routing protocols with IBS

5.1 Identifiers and public keys

Routing information secured by IBS should be based on IP addresses. These are the only well-known IDs at the network layer, and will be transmitted as part of the routing messages anyway. We believe the alternative of MAC addresses, as proposed in [3], represent a more static and vulnerable solution. There is no well-known approach for global or multi-hop dissemination of MAC addresses, and a hardware failure or change of node would demand a new key. However, with our suggestion, dynamic address assignment will be precluded, as fixed IP addresses will be needed.

5.2 Address configuration

Address requirements: The main OLSR address is a natural choice of IBS public key. It always follows as originator address in the OLSR routing messages. OLSR [7] assumes nodes are assigned addresses within a defined address sequence. This is not really needed for the OLSR protocol. However, the MANET nodes should be addressable from the external network. A gateway between the MANET and fixed network could hide the internal addresses. The Gateways could also be used to provide DNS service.

Assigning unique addresses: Several proposals regarding IP address assignment and how to resolve duplicate addresses in MANETs have been put forward. The suggestions implicitly assume that any node may join the MANET, and the IP addresses should be obtained as the node enters the network. In emergency and rescue operations, address administration could be co-located with the PKG (OSRM/RCC). The PKG can
thus prevent duplicate addresses. Merging nodes from multiple security domains call for international conventions on address assignment in Emergency and Rescue organizations. A fall-back is stateful address configuration by a PKG service, accessible through the gateway(s).

5.3 Securing OLSR with IBS

**OLSR protection schemes:** The protection schemes proposed for the OLSR protocol, [1][8][19], rely on signed routing messages and sequence numbers or timestamps. The messages are signed hop-by-hop or end-to-end. With end-to-end signatures, the OLSR main address is the natural choice of public key. It is forwarded unchanged end-to-end in the originator address field of the OLSR message. The hop-by-hop approach requires an ID of the retransmitting node as public key, e.g. the source IP address in the IP header.

**Sketch of protocol operation:** Micro mobility within the MANET is handled by the MANET routing protocol. Ideally, all MANET routing messages are signed. The routing messages are processed according to the specifications in [7]. The IBS signature is verified as the final step before the message is accepted as valid. In case OLSR is extended with multicast-tree building information, these messages should be signed and treated as any other routing messages.

Macro mobility could be managed by Mobile IP [13] (MIP) as described in [4]. Our assumption of fixed IP addresses maps with the MIP assumption of fixed home addresses. The MANET Gateways act as foreign agents (FAs) and provide temporary care-of-addresses (CoAs) reflecting the current internet attachment point of the mobile nodes (MNs). The CoA must be registered with a home agent (HA) router on the MN’s home network. The HA intercepts datagrams destined for the MN and tunnels them to the CoA. At the end of the tunnel the FA de-capsulates datagrams and forward them to the mobile node. Inside the MANET, the nodes may use their home addresses if these are not exposed to external networks. With the proactive routing protocol the nodes will know whether the destination is accessible within the MANET or the datagram should be routed to the FA Gateway. Seen from the fixed network, each FA gateway should possess different subnet IDs on the MANET side. This enables tunneling of IP-datagrams from the HA independently of MANET fragmentation and merging. The
subnet IDs need not be visible within the MANET. The FA should be set as default gateway. With more Gateways, the MANET node must choose which one to sign up with. Changes in connectivity may require a change in FA binding. The HA and previous FA gateway should be informed as usual in MIP. On splits, MANET to MANET communication can continue through the Gateways and MIP. We have assumed IPv4. IPv6 could be treated as implied in [15].

5.4 Performance

Assumptions and analytical approximations: The performance evaluation assumes a generalized protection scheme; OLSR messages extended with a signature field. OLSR message sizes are given in [2]. The methods for calculation of overhead and channel utilization are adopted from [18]. Our calculations are based on network topology information from [20]†. HNA and MID messages are assumed to contribute little to the overall OLSR traffic, and are therefore left out of the calculations. No message aggregation is assumed. TC intervals are set to 2,5*HELLO intervals, i.e. a HELLO interval of 2 seconds imply 5 seconds between each TC message.

There are few IBS implementations commercially available. In order to provide an estimate for processing delays and signature sizes, we assume Shamir’s original IBS scheme. Signing and verification is then comparable to RSA [14] operations, and implies a signature can be processed in a few milliseconds (Broadcom BCM5825 security processor provides a 1024bit RSA operation in 0,083ms.) In Shamir’s IBS scheme both signing and verification include a modular exponentiation with a large prime (e in Fig.1). Signing and verification cost is therefore similar. A modulus of 1024 bits gives a 2048 bits signature.

Processor utilization: With a probability, pMPR, of being selected as MPR, n nodes in the network and N 1-hop neighbors, the number of signatures every node must process per second equals

\[
(p_{MPR}*n*TC_{rate} + N*HELLO_{rate})
\]  (1)

Fig.2a shows the processor utilization for various signature processing times and

† Obtained through calculations and simulations with ns2 simulator. The number of MPRs selected by each node was set 10-20% higher than minimum required by the standard protocol. Other simulation parameters: Random Waypoint mobility model, maximum speed 20 m/s and pause time 60s, nodes distributed over an 1500mx300m area, transmission range 250m.

126
network sizes, given a HELLO interval of 2 seconds. Experience has showed that the default HELLO interval of 2s can be too long to handle mobile nodes. Fig. 2b) displays processor utilization with 32 nodes in the network for various HELLO intervals and signature processing delays. Fig. 2a) and b) illustrate that with a dedicated processor and processing delays in the order of a few milliseconds, processor utilization is a not limiting factor for the applicability of IBS.

**Overhead:** The size of the TC messages is calculated under the assumptions that every TC message advertises \((N-1)\times m_{S RATE}\) links \((N\) equals the number of 1-hop neighbors and \(m_{S RATE}\) the MPR selector rate; the average number of nodes that has selected this node as its MPR). The HELLO messages are assumed to include two different link codes. Fig.3a) displays the average OLSR traffic imposed per channel for various network sizes and signature lengths, given a HELLO interval of 2s. The figure indicates a maximum network size around 32 nodes under the assumption of 2048 bit signatures and a requirement that imposed OLSR traffic should not exceed 10% of the 1Mbps bandwidth. Fig.3b) displays the consequences of various HELLO intervals and signature sizes with a network size of 32 nodes as example. It shows that the overhead introduced by OLSR signatures restrain shorter message emission intervals, especially with signature sizes larger than a kilobit.

**Channel Utilization:** The channel utilization is measured as average fraction of a second a channel is occupied. The calculations assume IEEE 802.11b with CSMA/CS-DSSS for the lower layers. OLSR traffic is broadcast traffic with a data rate of 1Mbps. UDP, IP and 802.11 MAC headers add an overhead of 8+20+34 bytes to each OLSR packet, respectively. A collision and error free channel is assumed. Additional 802.11b delay is 552us [9] per MAC frame. Fig. 4a) shows the average channel occupation for various network and signature sizes with a HELLO interval of 2s. Assuming that OLSR traffic should not utilize more than 10% of the channel capacity, the figure shows that without signatures, the maximum network size is limited to 42 nodes. A signature of 2048 bits reduces the maximum network size to approximately 26 nodes. Fig. 4b) shows the channel utilization for various signature sizes and HELLO intervals in a network of 32 nodes. It shows that channel utilization confines possible OLSR message emission rates, especially with signatures of more than some hundred bits.

**Discussion:** The calculations show that whereas processing delays of some milliseconds have little impact on the applicability, signature sizes in the order of thousands of bits have. This is not limited to IBS schemes. Channel utilization confines
the maximum size of the network and possible OLSR message emission rates. Shamir’s IBS doubles the size of the signatures compared to RSA for the same modulus. Shorter signatures are desirable. Recent IBS schemes [6] [12] [17] rely on bilinear pairings. Super-singular elliptic curves imply signature sizes comparable to RSA signatures. (Short signatures from the pairings are proposed in [5], but these are not identity-based.)

Our calculations assumed a single verification per message and duplicate messages discarded. More attempts may be needed if the first verification fails. Hence, processing delays can be a more limiting factor than our calculations indicates. However, other IBS schemes generally have a different computational cost for signing and verification. Each node verifies more signatures than it signs. Longer signature generation times may be acceptable.

**Fig. 2** Processor utilization versus processing delays for:

- **a)** various network sizes, HELLO interval=2 seconds (n = number of nodes),
- **b)** various HELLO intervals, 32 nodes

**Fig. 3** OLSR traffic imposed per channel for various signature sizes with:

- **a)** various network sizes, HELLO interval =2seconds,
- **b)** various HELLO intervals, 32 nodes
6 Related work

Khalili, Katz and Arbaugh [10] suggest identity-based cryptography for ad hoc networks, focusing on key distribution. They suggest spreading the private key generator over more nodes with the aid of threshold cryptography. However, they do not analyze the network implications of identity-based schemes.

The IETF Host Identity Protocol (HIP) [11] suggests a new namespace that decouples the internetworking layer from the upper layers. With HIP, the host identifier, not the IP address, represents a long term identifier.

7 Conclusions and future work

This analysis of identity-based signature (IBS) schemes for protection of MANET routing information revealed channel utilization (assuming IEEE 802.11b) as the most limiting factor. The signature sizes confine the network size and possible routing message emission frequencies. Signatures of more than a few hundred bits are only applicable in small MANETs with standard message emission intervals. This result is not limited to identity-based signatures. From an addressing and key distribution point
of view, IBS fit emergency and rescue operations involving participants from a single security domain. Operations with participants from multiple security domains require distribution of new parameters in the merging networks.

Our analysis assumed the OLSR routing protocol. We believe the results have relevance also for other MANET routing protocols. However, reactive protocols have a different message generation pattern. Further simulations and experiments are needed to verify the concept of IBS protected MANET routing protocols combined with Mobile IP. Increased bandwidth efficiency is desirable. Hybrid solutions are topics for further work. Alternatively, messages could be signed only when changes are detected. Another topic for further studies is evaluation of the cost/gain of various MANET routing protocol protection schemes.

References


Corrections:

- In the OLSR section it is claimed that routes are calculated from tc messages—should have been tc and hello messages.
- Topology change—should have been topology control
PAPER III: Hybrid protection of OLSR

Anne Marie Hegland, Pål Spilling, Leif Nilsen, and Øivind Kure

To appear in Proceedings of
Workshop on Cryptography for Ad Hoc Networks
(WCAN’06)
held in conjunction with
the 33rd International Colloquium on Automata, Languages and Programming
(ICALP 2006),
Venice, Italy, July 16th, 2006
Hybrid Protection of OLSR

ANNE MARIE HEGLAND* PÅL SPILLING* LEIF NILSEN* ØIVIND KURE†

*UniK – University Graduate Center at Kjeller
P.O. Box 70, NO-2027 Kjeller, Norway
†NTNU,Trondheim, Norway
{amh;paal;leif.nilsen;okure }@unik.no

Abstract
The network layer of ad hoc networks is prone to attacks. Insertion of false routing information is inherently easy unless the routing messages are authenticated. Hence, protection of the routing protocol is a prerequisite for a reliable network service. The wireless broadcast media calls for bandwidth efficient solutions. This paper proposes a hybrid protection scheme for the ad hoc Optimized Link State Routing protocol (OLSR). Identity-based signatures are combined with hash chains. Only some messages are signed. Unsigned messages include a hitherto undisclosed value from the hash chain as a lightweight proof of authenticity. The result is a bandwidth efficient scheme providing adequate protection.

Key words: Secure routing, ad hoc, authentication, hash chains, identity-based signatures, performance

1 Introduction

Protection of the network layer in mobile ad hoc networks calls for unilateral authentication of the routing messages. Possible solutions include symmetric message authentication codes, traditional digital signatures and identity-based signatures.

Symmetric schemes are efficient both regarding computational efficiency and bandwidth consumption. However, with a single group key, the security succumbs to a
single captured node. Source authentication is precluded, and malicious insiders may masquerade as others. Furthermore, exclusion of compromised nodes is hard. The alternative of pairwise exclusive keys provides better intrusion tolerance, but scales badly.

Asymmetric (public key) schemes enable revocation of single nodes and provide better protection against Byzantine behavior from insiders. A drawback of traditional digital signature schemes is the potential need for bandwidth consuming certificate exchanges. The identity-based signature (IBS) schemes make certificates obsolete. If information such as IP-addresses that are by default transferred in the routing messages is used as public keys, identity-based schemes may scale better than the certificate based schemes.

In [7] it is shown how the signature sizes confines the maximum size of the network and possible routing message emission frequencies. Bandwidth efficient solutions were sought. This paper proposes a hybrid protection scheme for proactive link state routing protocols, and provides specific formats for its implementation with the optimized link state routing protocol (OLSR) [3]. The aim is a bandwidth efficient solution providing an adequate level of security. Both the security and the performance of the protection scheme are analyzed.

The hybrid protection scheme combines asymmetric, identity-based signatures with shorter and computationally cheaper hash values from a hash chain [11]. Bandwidth is saved by signing only some routing messages. Unsigned messages include a hitherto undisclosed hash value from the chain. The values serve as lightweight proofs of authenticity, and make it hard for malicious nodes to successfully insert additional and false routing messages on other nodes’ behalf, even if not all messages are signed. The hash values also ease sequence number wrap around detection and protect against replays.

The rest of this paper is organized as follows. OLSR is described in section 2 and threats are described in section 3. Section 4 outlines the basic security mechanisms. In section 5 the hybrid protection scheme is described and analyzed. The performance of the protection scheme is evaluated in section 6. Related work is described in section 7. Conclusions and future work are discussed in section 8.
2 The Optimized link state routing protocol (OLSR)

OLSR [3] is a proactive link state routing protocol designed for large and dense mobile ad hoc networks. The packet and message formats are shown in Fig.1. Bandwidth consumption is optimized compared to classical flooding through the use of Multipoint Relay (MPR) nodes. Only nodes that are chosen as MPRs forward routing information. Each node selects MPRs from its 1-hop neighbors in such a way that all 2-hop neighbors are covered by at least one MPR node.

![Fig. 1. OLSR packet and message formats](image)

Each node broadcasts its links and MPR selections through HELLO messages. HELLO messages are not forwarded. Link state information is disseminated through Topology Change (TC) messages that are generated and flooded in the entire network by the MPR nodes. Routing tables are computed from the information exchanged through the TC and HELLO messages. The 1-hop and 2-hop neighbors are detected from the HELLO messages. Routing information concerning 3-hop nodes and more distant nodes are found from the TC messages.

An OLSR node may have several network interfaces, but there is always one main...
address identifying the node. The main address is used as the originator address in all OLSR messages [3]. Other OLSR interfaces are announced through *Multiple Interface Declaration* (MID) messages. These are visible as source and destination addresses in the IP header. OLSR packets are encapsulated in UDP over IP. Interfaces to non-OLSR networks are declared through *Host and Network Association* (HNA) messages. Like TC messages, MID and HNA messages are flooded by the MPRs.

3 Threats

The OLSR routing protocol is vulnerable to a number of attacks that can be classified as Denial of Service (DoS) attacks. The attacks may aim at a single node, a group of nodes or the whole MANET. The threats originate both from external malicious nodes as well as internal, legitimate members of the network not behaving according to protocol. Possible attacks include:

*Exhaustion*: Insertion of routing information for the purpose of exhausting other nodes or congesting the channel.

*Route corruption*: Insertion of bogus routing information or modified routing messages in order to corrupt the routing tables.

*Replay*: Retransmission of earlier routing messages to corrupt routes or exhaust other nodes.

*Blackhole* nodes attract traffic by falsely advertising they have the best routes to others. The target may be control of the traffic flow. Blackholes may choose to forward routing information but not payload.

*Masquerade*, i.e. nodes pretending to be others, can be launch by replays of old routing messages or insertion of bogus routing information. The aim can be route corruption or node exhaustion.

*Selfish nodes* refuse to forward or only selectively forward routing messages from others and give priority to own traffic.

*Wormholes*, set up by colluding nodes tunneling routing messages through a high speed link from one part of the network to another, can make geographically distant nodes erroneously believe they are neighbors. The colluding nodes may act as Blackholes.

*Eavesdropping* is assumed not to disturb routing. However, eavesdropping can be
used for traffic analysis and monitoring of network topology. The information learned could be utilized as background information for other attacks.

*Byzantine behavior* refers to nodes that do not behave according to the protocol. This includes insertion of false or modified routing messages, masquerade, selfish behavior, delayed response and similar deviations from the protocol.

### 4 Basic Security Mechanisms

**Identity-based signatures (IBS):** *Identity-based Cryptography* [18] removes the need for certificates whereas preserving the benefits of public key systems. IDs such as IP-addresses that are by default transferred in the routing messages can be used as public keys. The first identity-based signature (IBS) scheme, introduced by Shamir [18], was derived from RSA [16]. More recent schemes such as [4],[8],[14], and [17] are based on bilinear pairings.

During the setup phase of an identity-based scheme, the trusted entity, the Private Key Generator (PKG), chooses a secret master key and generates the corresponding public system parameters. Afterwards, in the extraction phase, it issues private keys. The private keys are calculated by combining the IDs and the private master-key of the PKG. Signatures, generated by the private key, are verified with the aid of the sender ID plus the public system parameters.

**Hash chains** [11] are constructed by repeated hashes of an initial random seed \( RND \):

\[
h_1 = h(RND), \quad h_2 = h(h_1), \ldots, h_{n-1} = h(h_{n-2}), \quad h_n = h(h_{n-1}).
\]

The hash function \( h() \) is a cryptographic one-way function. The last value in the chain, i.e. the hash anchor \( h_n \), is distributed through an authenticated channel. This commitment enables its holder to verify whether later disclosed values originate from the same hash chain or not. Repeated hashes of the received values should return the hash anchor. Forward hash calculation is fast and easy. However, the one-way property makes it hard for anyone but the creator to find the preceding values in the hash chain.
5 A hybrid Protection Scheme for OLSR

5.1 Aim and assumptions

The aim of the hybrid protection scheme is a reliable network service. The scheme seeks to balance bandwidth efficiency and security. The routing messages are signed periodically. TC messages are signed when topology changes, or periodically if no changes are detected. In addition, all routing messages include a previously undisclosed value of a hash chain. The hash values make it hard for malicious nodes to succeed in routing protocol attacks even if not all messages are signed.

The hybrid protection scheme assumes an IBS scheme where the OLSR main addresses are used as public keys, and the corresponding private keys and system parameters have been derived by a separate key management service. This implies a planned ad hoc network.

5.2 Sketch of protocol operation

Message formats: The message formats are shown in Fig.1. The hybrid protection scheme is based on the message sequence number (MSN) receipt technique for replay protection suggested in [20], i.e. the HELLO messages include not only the address of their neighbors but also the last MSN received from them. Ref. [20] suggests the MSN field is extended from 16 to 32 bits to avoid MSN wrap-around during network operation. The hash chain removes the need for more bits in the sequence number field. The hash value will prove whether a message with a lower MSN than previously received is fresh or not.

If all messages were either hashed or signed and hashed, and the signature and hash sizes are fixed, one bit would be enough to signal signature type to the receivers. A more general format is suggested with the SType and signature & hash size fields shown in Fig.1. The SType field can be used to signal signature type plus configuration parameters such as signature and hash algorithms and the maximum interval between signed messages.

Hash anchor distribution: The first hash value received in a signed message is stored in the receiver as the hash anchor of that originator. Note that this value does not have to be the original anchor of the chain. The hash value in every signed message represents a refreshed hash anchor.
MSN and the length of the hash chains: The hash chains are made long enough to avoid initialization of new chains during network lifetime. This is both to ease MSN wrap-around recognition and to increase the robustness to packet losses. If a signed message is lost, the receiver can still verify the hash value of the next message received.

Whereas a single hash operation is computationally cheap, long hash chains imply more processing in the sender node. That is to “spool” from the seed to the wanted value when a new message is prepared. A trade off between work load and storage space is made by storing some intermediate values of the hash chain. The sender then only has to “spool” from the closest stored value in the chain.

Routing message processing: When a signed message is received and the signature check fails, the message is discarded. Otherwise, the receiver stores the hash anchor and the message sequence number (MSN). When hashed messages are received, the node verifies that repeated hashes of the received value return the hash anchor. Subtracting the stored MSN number from the received MSN value provides the number of repeated hashes required.

1) Local messages: When a new 1-hop neighbor is detected, but it has not yet been authenticated through a signed HELLO message, the neighbor and the link are set to “non-verified”. Still, if the hash value can be verified, i.e. a hash anchor has been received through flooded messages or the node has been in the neighborhood recently, the neighbor is accepted as a symmetric neighbor. In case no hash anchor is available, the link status is set to asymmetric until the link has been authenticated through a correctly signed HELLO message. MPRs are chosen among the authenticated nodes only, and only authenticated nodes are accepted as MPR selectors.

2) Flooded messages are accepted over symmetric links only. The MPRs only flood messages from their authenticated MPR selectors, and only authenticated neighbors are advertised in the TC messages. TC messages are flooded regularly. TC messages announcing topology changes are signed. The TC messages are also signed if no topology changes have been detected within a given period. If a topology change is detected in an unsigned hashed message (the signed message may have been lost), the new entry is marked “non-verified”. The status is changed to “verified” when a correctly signed message is received. Correctly signed messages take precedence over unsigned messages.

HNA and MID messages are signed periodically. The contents of these messages are expected to change little over time. Most messages will be used to refresh earlier
entries. When a new interface-to-main address association is listed in an unsigned hashed MID message, the new entry in the multiple interface association base is marked “non-verified”. When a corresponding entry is received in a signed message, the status is changed to “verified”. In case of conflicts, verified entries take precedence over non-verified. HNA messages are treated similarly. Verified entries are preferred as gateways.

5.3 Security analysis

The cryptographically signed routing messages provide source authentication and proof of integrity. They also provide fresh, authenticated hash anchors. The hash values do not guarantee the integrity of the messages, but make successful attacks harder then when the messages are left unsigned. The receivers expect fresh values from the hash chains. Else, the messages are discarded. The hash values make it impossible for malicious nodes to insert additional messages on other’s behalf. The malicious nodes must wait for the true originator to disclose the next value in the chain. In addition, modified messages that arrive later than the original message are discarded. Hence, the hash values make successfully timed attacks hard.

Local messages: Ideally, a node should not set the link status to symmetric before the link has been authenticated through a signed HELLO message as the routing tables are recalculated on detection of new 2-hop neighbors. However, accepting hash values as a temporary proof of trustworthiness makes neighbor detection faster at the same time as it is not straight forward for an attacker to launch an attack. The MPR selection is more important to protect as these nodes act on behalf of others. This is accounted for as only authenticated neighbors are accepted as MPRs and MPR selectors.

Some attack situations for local messages are shown in Fig.2 a).

Situation 1: The true originator node A is within 1-hop neighborhood of the receiving node B. If the packet is not lost, the receiver will receive the message from the true originator at the same time as any malicious node. Any replay or modified messages received from the malicious node will arrive later than the true message, and be discarded.

If the packet is lost on the way from A to B, the malicious node C may retransmit a modified message. If the message is signed, any modification will be detected by B. If it was left unchanged, node C basically improves the connectivity. On the other hand, if
the message was unsigned, node C may succeed in introducing false information. That is, it requires C to spoof the interface address of A, otherwise node B will detect that this packet was received from a different OLSR interface. This spoofing is detectable by node A. In addition, the next message received directly from A will remove any false information introduced by C. Both with and without packet losses, the hash values prevent C from introducing additional and unscheduled HELLO message on A’s behalf.

**Situation 2:** Node C replays HELLO messages from node A recorded earlier in another part of the network. If node B already has a fresher hash anchor obtained through flooded message from A, the replayed message is discarded. Furthermore, neither Byzantine insiders nor external attackers will be able to produce a signed message from A including a fresh MSN-receipt from B, which is required to establish a verified symmetric link.

**Situation 3:** Colluding nodes C1 and C2 forward traffic between A and B through a high speed link in a wormhole attack. If C1 and C2 retransmit the messages unchanged, they may improve network connectivity. Still, they could also act as Blackholes and forward routing information, but not payload.

Signatures are not enough to protect against wormholes. Precise time information or location awareness or similar is required in order to mitigate such attacks. Mobility also helps. We believe wormhole attacks are hard to mount in ad hoc networks. Their importance should not be exaggerated.

**Situation 4:** Node A is out of 1-hop neighborhood of node B. A malicious node C is placed between node A and node B. The malicious node may replay the HELLO messages from A unchanged or forward modified versions. This is the same as a wormhole attack, except that node C represents both ends of the “tunnel”. Unless C retransmits the messages on orthogonal OLSR interfaces, nodes A and B will hear the retransmission by C. The hash values prevent C from inserting additional and unscheduled messages on A and B’s behalf. Signed messages erase any false information inserted by C.

**Situation 5:** A malicious node inserts false messages. The signature and hash validations will fail.
Situation 1: B and C both within 1-hop neighborhood of A

No packet loss  Packet loss

Situation 2: C moves from 1-hop neighborhood of A to 1-hop neighborhood of B. A and B not neighbors

Situation 3: Wormhole C 1-hop neighbor of A and B A and B not 1-hop neighbors

No packet loss  Packet loss

Situation 4: C 1-hop neighbor of A and B, A and B not 1-hop neighbors

Situation 5: C inserts false information

Situation 1: B and C both within 1-hop neighborhood of the MPR (A)

No packet loss  Packet loss

Situation 2: MPR (B) and C 1-hop neighbors of A

Situation 3: Shorter route through C

Situation 4: C between MPR B and the MPR selector A

Situation 5: C multiple hops from originating MPR

Situation 1&2: There is a malicious node C within the 1-hop neighborhood of the neighbors A and B. This is parallel to situation 1 for local messages. The hash values prevent C from successfully inserting additional and unscheduled messages on the behalf of A (and B). With no packet losses, B will receive the correct message from A at the same time as C, and discard any later modified or unmodified duplicates. In order to make B flood false information from A it requires the packet was lost by B, and B is an MPR of node A, and C spoofs A’s address. Such a masquerade can be detected, but not prevented by A. However, more than one packet must be lost in sequence in order to
make B flood false topology information. This is as B expects TC messages to be signed upon topology changes. It also requires A is a MPR (only MPRs generate TC messages), and B must be its MPR.

*Situation 3&4:* In situation 3, the malicious node C represents a shorter route from MPR A to the receiver, and forward the routing message to node B before the message is received through an authenticated route. In situation 4 there is a malicious node between A and its MPR B. Situation 3 and 4 are basically wormhole attacks where node C represents both ends of the “tunnel”. C may succeed in temporarily introducing false information. Mobile nodes make it harder for node C to succeed in a persisting attack. It requires C to track both A and B and maintain the shorter route in situation 3 or keep A and B strict 2-hop neighbors in situation 4. The hash values inhibit C in introducing new messages from A before A has released the next value in the hash chain. Signed messages erase any false information introduced by C.

*Situation 5:* Node C is two or more hops away from the originating MPR A. When C receives a new hashed message; C can modify the contents and retransmit the message. Still, it requires C to spoof the address of a symmetric neighbor of node B. In addition, C must send the modified packet before the true MPR forwards the (correct) message, as B will discard later duplicates. The hash values make it impossible for C to insert a false packet on behalf of the MPR before a new value in the chain has been received.

**Concluding remarks:** Evaluating the hash values before the computationally more expensive signature verification renders the hybrid scheme more robust to node exhaustion attacks than schemes solely relying on asymmetric signatures. Knowledge of the routing message emission rates and the reception time of the last authenticated MSN can be used judge whether the received MSN is reasonable or not. This stops possible hash calculation exhaustion attacks.

Neither signatures nor hash values preclude selfish nodes. It is hard to judge whether an MPR failed to forward a packet due to selfishness or for other reasons.

### 6 Performance

#### 6.1 General assumptions and analytical approximations

The network topology information and method for calculation of channel utilization are adopted from [19]. The network topology information in [19] was obtained through
calculations and simulations with the ns-2 network simulator. The number of MPRs selected by each node was set 10-20% higher than the minimum required by the standard protocol. Other simulation parameters were: Random Waypoint mobility model, maximum speed 20 m/s and pause time 60s, nodes distributed over 1500mx300m, and transmission range 250m.

The performance calculations are based on the packet formats of the hybrid scheme and standard OLSR shown in Fig.1. HNA and MID messages are assumed to contribute little to the overall OLSR traffic, and are therefore not included in the calculations. No message aggregation is assumed, and the TC intervals are set to 2,5*HELLO intervals. I.e., a HELLO interval of 2 seconds implies 5 seconds between each TC message.

Whereas processor utilization is important, the analysis in [7] revealed channel utilization as a more limiting factor for the applicability of asymmetric signatures for protection of ad hoc routing messages. The replacement of some of the asymmetric signatures with cheaper hash functions makes processor utilization even less constraining. The focus of this paper is therefore on channel utilization, and processor utilization is left out of further discussions.

The performance evaluation assumes asymmetric schemes and hash functions of similar security levels, and is based on the parameter sizes proposed by NIST [12], shown in A.1 in Appendix A. Corresponding signature sizes are found in table A.2 in Appendix A.

Shamir’s IBS expands the signature size with a factor of two compared to RSA for the same modulus. Depending on implementation, the newer IBS schemes from the pairings such as [4], [8], [14], and [17] can provide shorter signatures. In [12], the proposed hash sizes are twice the size of the symmetric keys. This is for collision resistance. The threat to the hash chains in the hybrid scheme is inversion of the one-way function (preimage resistance) rather than collision resistance. Consequently, the hash sizes can be halved.

6.2 Channel utilization

The channel utilization is measured as the average fraction of a second a channel is occupied. The calculations assume IEEE 802.11b with CSMA/CS-DSSS for the lower layers. OLSR traffic is broadcast traffic with a data rate of 1Mbps. UDP, IP and 802.11 MAC headers add an overhead of 8+20+34 bytes to each OLSR packet, respectively. In addition comes a 802.11b delay of 552us [10] per MAC frame. A collision and error free channel is assumed.
The size of the TC messages is calculated under the assumptions that every TC message advertises \((N-1)\times m_{SRATE}\) links \((N\) equals the number of nodes within 1-hop neighborhood and \(m_{SRATE}\) the MPR selector rate; the average probability that a neighbor selects this node as its MPR). The HELLO messages are assumed to include two different link codes.

Fig.3 shows the average channel utilization measured in fraction of a second occupied by OLSR traffic for various network sizes, security overheads and signature to hash ratios. Fig. 3a) shows that most of the bandwidth savings is obtained by the first hashed messages. Increasing the number of hashed messages to each signed message is most beneficial for the largest signatures.

![Average channel occupation by OLSR](image)

*a) various network sizes and signature overheads*

![Average OLSR channel occupation](image)

*b) different signature to hash ratios*

*Fig. 3. Average channel utilization of the hybrid protection scheme*
Whereas Fig. 3a) assumes the same signature to hash ratio for all message types, Fig. 3b) includes the effect of differentiation on message types. The notation “2048/112: H(1:3), TC(1:0)” means every fourth HELLO messages include a signature of 2048 bits and a hash value of 112 bits. All TC messages include a signature of 2048 bits, but no hash value. Likewise, 1024/0:(1:0) refers to a pure signature scheme with 1024 bits signatures added to both the HELLO and the TC messages. The figure shows that most of the bandwidth gain is obtained through hashed TC messages rather than hashed HELLOS.

6.3 Discussions

As a rule of thumb, OLSR traffic should not exceed 10% of the bandwidth. The calculations indicate that significant bandwidth savings can be obtained by introducing one or more hashed messages for every signed one. More hashed messages per signed message are especially beneficial with the larger signature sizes, more nodes in the network and increasing emission frequencies. With increasing network sizes, the average number of TC messages grows faster than the average number of HELLO messages per channel, and the calculations show that the signed TC messages contribute more to the channel utilization than the signed HELLO messages. As can be seen from Fig. 3b), this difference is amplified with increasing network sizes. Through calculations, we found that the difference is also augmented with shorter routing message emission intervals. (Experience has shown that the default HELLO interval of 2s can be too long with mobile nodes). Hence, hashing the flooded messages is desirable. Regarding local messages, under the assumption that only authenticated neighbors are accepted as MPRs and MPR selectors, the gain obtained by hashing the HELLO messages should be weighed against the potential added delay in the neighbor detection and MPR selection processes.

Whereas TC message signatures should be triggered by topology changes or the maximum time since last signature is exceeded, the calculations assume periodical (maximum time triggered) signatures. Further simulations are required to confirm the average time between topology changes.

The calculations here only considered average channel occupation. That is, (2048/80:(1:3)) would show the same average channel occupation as (1024/80:(1:1)). However, the probability of congestion and packet losses may be higher with the larger
signature sizes. Generally, shorter signatures are preferred. Furthermore, the calculations assumed an IBS scheme. But the figures are applicable to any signature scheme with the chosen signature sizes. If the IBS scheme is replaced by a traditional digital signature scheme, the bandwidth consumption of certificate exchanges must be taken into consideration. An estimate is found by adding the certificate lengths to the signature sizes.

7 Related work

SOLSR [6], proposed by Hong, Hong, and Fu, uses asymmetric signatures and hash chains for the protection of OLSR. However, the SOLSR approach is different from the hybrid protection scheme. Similarly to the hybrid protection scheme, the signatures in SOLSR cover all non-mutable fields of the routing messages. But SOLSR assumes all routing messages include a digital signature. Bandwidth efficiency is not considered. The hash chains are used to protect the mutable Time to live (TTL) and HopCount fields, and do not serve as a proof of authenticity of unsigned messages as in the hybrid protection scheme. Each new SOLSR routing message contains both the seed and the anchor of a new hash chain. The hash anchor is included in the signed part of the routing message. The seed is updated by intermediate nodes. Hashing the seed received a number of times equal to the difference between the TTL and the HopCount should return the hash anchor. The hybrid protection scheme leaves the TTL and HopCount fields unprotected.

In our opinion, there are some deficiencies with the SOLSR scheme. According to the OLSR RFC [3]; intermediate nodes should decrease the TTL and increase the HopCount fields nodes before the routing message is forwarded. However, whereas the hop count plays an important role in the calculations of the shortest routes in distance vector routing protocols such as AODV [15], there is basically no need to protect the TTL and HopCount fields of OLSR. The HopCounts are not used in the route calculations. Reducing the TTL value by more than one means the packet will die sooner. Still, any node can decide not to forward the packet, which kills it even faster. Increasing the TTL will not increase the packet lifetime significantly. Duplicates of flooded packets are detected on the basis of sequence numbers, and will be discarded. Furthermore, SOLSR detects wormholes on the basis of observed round-trip times. A
wormhole is suspected if the round-trip time indicates the distance to its neighbor exceeds the maximum transmission range. However, it is not evident how one can judge whether an extra delay is caused by lower layers normal media contention or an additional travel distance through a wormhole tunnel.

The reactive SAODV [21] uses hash chains to authenticate the hop count, which is used to calculate the shortest routes. In SLSP [13], hash chains are used to prevent topology information from being distributed further than a specified number of hops.

In SEAD [9], a hash chain is used to ensure other nodes cannot retransmit the routing message with a higher sequence number than just received or with a shorter distance than currently received. Loose time synchronization is required in order to prevent successful masquerade attacks with the use of already disclosed hash values. In our scheme, hash chains are combined with signatures and message sequence number receipts and periodic proactive refreshes of the hash anchor (signed messages). This reduces the need for time synchronization.

Whereas the hybrid protection scheme is based on the MSN-receipt technique in [20], Adjih et al. [1] and Hafslund et al. [5] propose time stamps instead of sequence numbers for replay protection. In [19] it is shown in that the MSN-receipt technique scales better than the time stamp solutions.

8 Conclusions and future work

Bandwidth efficient security solutions are sought for ad hoc networks. The hybrid protection scheme shows how hash values from a hash chain can be combined with asymmetric identity-based signatures (IBS) into a bandwidth efficient security solution. The calculations showed that significant bandwidth savings can be obtained even by in average hashing only every second routing message instead of signing it. More hashed messages increase the savings.

The hybrid protection scheme mitigates persisting route corruption, exhaustion, masquerade, replay and black hole attacks from both externals as well as (to some extent) internal legitimate members of the network. Although not all routing messages are signed, the broadcast nature of the routing protocol combined with the hash values make successful attacks harder. The hash values also eases message sequence number (MSN) wrap-around recognition and make replay attacks easier to detect. The hash
values prevent attackers from successfully inserting unscheduled routing messages on the behalf of other nodes. However, under specific circumstances, as discussed in the security analysis, there is a risk that malicious nodes may succeed in temporarily introducing false routing information. These attacks are not straightforward to mount, and signed routing messages will erase any false information. TC messages are expected to be signed if they include a topology change. It would thus require two or more lost packets in sequence in order to insert false topology information. A topology change from one hashed message to the next is suspicious.

Whereas the hybrid protection scheme is proposed for OLSR, we regard the method as generic and applicable to other proactive link-state routing protocols as well. Furthermore, the IBS scheme could be replaced with a digital signature scheme. Certificate distribution must then be taken into consideration. Bandwidth efficient protection schemes for other routing protocols are topics for further research.

The security levels suggested in [12] were used as an estimate for the performance evaluation. However, a comprehensive framework for the decision of appropriate key lengths in ad hoc networks is a topic for further investigations. So is effective key management, including revocation.

References


A Appendix: Security levels, hash lengths and signature sizes

(From NIST [12])

<table>
<thead>
<tr>
<th>Security level (bits)</th>
<th>Hash size</th>
<th>IFC &amp; TDL ( n )</th>
<th>FFC (SDL) ( p/q )</th>
<th>ECC key length</th>
<th>Safe until year</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>160</td>
<td>1024</td>
<td>1024/160</td>
<td>160-223</td>
<td>2010</td>
</tr>
<tr>
<td>112</td>
<td>224</td>
<td>2048</td>
<td>2048/224</td>
<td>224-255</td>
<td>2030</td>
</tr>
</tbody>
</table>

Table A.1
Comparable security levels

<table>
<thead>
<tr>
<th>Security level</th>
<th>Hash size</th>
<th>IFC &amp; DL</th>
<th>SDL</th>
<th>ECC</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>80</td>
<td>RSA</td>
<td>DSA</td>
<td>ECDSA</td>
</tr>
<tr>
<td>112</td>
<td>112</td>
<td>Shamir</td>
<td></td>
<td>BLS[2]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Security level</th>
<th>Hash size</th>
<th>IFC &amp; DL</th>
<th>SDL</th>
<th>ECC</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>80</td>
<td>RSA</td>
<td>DSA</td>
<td>ECDSA</td>
</tr>
<tr>
<td>112</td>
<td>112</td>
<td>Shamir</td>
<td></td>
<td>BLS[2]</td>
</tr>
</tbody>
</table>

Table A.2
Hash lengths and signature sizes

*) Depends on the embedding degree \( k \) of the torsion group \((1 \leq k \leq 6)\).

IFC= integer factorization cryptography
TDL= Traditional discrete log
SDL= subgroup discrete log
ECC= elliptic curve cryptography
\( n \)= bits in modulus
\( p \)= bits in prime modulus
\( q \)= subgroup size in bits
IBS= Identity-based signatures
Cert.= Certificate
Corrections:

- In the OLSR section Topology change – should have been topology control.
- In the paragraph describing SEAD in the Related Work section, it is claimed that loose time synchronization is required in order to prevent successful masquerade attacks with the aid of already disclosed hash values. This refers to the hash values of the TESLA authentication scheme. TESLA is one of more options that can be used to authenticate the source of each routing update messages. The text may give the impression that the time synchronization is required directly in the hash chains used to protect against short distance frauds and sequence number deceptions.
PAPER IV: On the Distribution of Revocation Information in Ad Hoc Networks

Anne Marie Hegland, Pål Spilling, Øivind Kure, and Leif Nilsen

To appear in Proceedings of the Third International Conference on Wireless and Mobile Communications (ICWMC’07)
Guadeloupe, French Caribbean, March 4-9, 2007
Abstract—This paper proposes a simple, scalable and robust scheme for the
distribution of revocation information in mobile ad hoc networks (MANETs); the MRL
scheme. A MRL is a revocation list for a specific MANET instantiation. The scheme is
designed for revocation of keys used to protect the ad hoc network service, and the MRLs
are distributed with the routing messages. MRLs are established with the aid of trusted
gateways reporting MANET nodes to a central trusted entity. The trusted gateways must
be able to detect which nodes are in the MANET. This comes intrinsically with proactive
ad hoc routing protocols, and may be achieved at some additional cost with reactive
protocols. The scheme is intended for ad hoc networks with a planned origin, and where
a common point of trust exists.

Key words: Key Management, ad hoc networks, revocation, MRL, security

I Introduction

Protection of the network layer can be achieved with the aid of unilateral
authentication of the routing messages. Possible solutions include symmetric message
authentication codes and asymmetric (public key) schemes such as digital and identity-based [22] signatures. Symmetric schemes are efficient both regarding computational efficiency and bandwidth consumption. However, changes in group membership imply re-keying. This represents a threat to network availability. There is no guarantee that all nodes receive the new key in a timely manner. Although still challenging, the exclusion of specific nodes is easier with asymmetric schemes. A survey of key management schemes for the protection of routing information in ad hoc networks is found in [12].

In traditional public key schemes, a trusted entity signs certificates that bind public keys to IDs. In identity-based schemes, IDs, e.g., IP-addresses, serve as public keys. The corresponding private keys are derived by a trusted entity that knows the secret system key. In traditional public key schemes revocation refer to the invalidation of certificates. In identity-based schemes IDs are revoked.

This paper focuses on revocation of certificates and IDs used to protect routing information in MANETs intended for operational scenarios such as emergency and rescue operations. This means ad hoc networks with a planned origin, and where common points of trust exists, and pre-configuration is possible. Our MANET is connected with the internet through one or more gateways, a so-called hybrid MANET. Only authorized nodes are allowed to join the MANET. Exclusion of captured or malfunctioning nodes necessitates distribution of revocation information over the ad hoc network.

A number of methods for certificate revocation have been proposed [24] [25]. Unfortunately, these are generally not very well suited for the ad hoc environment. The online certificate status protocol (OCSP) [19] and the Certificate Revocation Status (CRS) [16] approach require online access to a third party to check the validity of certificates. Online access is hard to guarantee at all times in MANETs.

In the certificate revocation list (CRL) method a Certificate Authority (CA) periodically updates a signed and time-stamped list of all revoked certificates. A major drawback is that with certificate lifetimes typically measured in years, even a small revocation rate may lead to long lists. Furthermore, new lists every time a node is revoked lead to large update costs. Therefore, CRLs are typically updated only weekly or biweekly [25]. Higher granularity is needed in order to be able to expel nodes from MANETs that last hours or days.
Δ-CRLs are shorter and provide fresher information. These only list certificates that have been revoked since last CRL update. But Δ-CRLs are also issued with a frequency, e.g., daily, that may be too low. Higher frequencies may lead to a significant update cost. Besides, if one or more of the ad hoc nodes missed the last CRL, the CRL may need to be distributed in the MANET anyway. Moreover, CRLs and Δ-CRLs normally list revocations concerning the entire security domain and not only the subset of nodes that takes part in a specific MANET. This represents a waste of bandwidth.

A revocation scheme for the ad hoc environment is needed. The contribution of this paper is the MRL scheme for efficient distribution of revocation information in ad hoc networks.

The rest of this paper is organized as follows. Related work is described in section II. Section III lists the requirements for revocation schemes for ad hoc networks. The MRL scheme is described in section IV. Section V sketches implementation of the MRL scheme with various routing protocols. The scheme is analyzed in section VI. Conclusions and suggestions for future work are found in section VII.

II Related Work

Crépeau and Davis [6] propose an accusation-based revocation scheme for ad hoc networks. Certificates are issued by an off-line certificate authority prior to network participation. All nodes monitor the behavior of the others. Accusations are posted if discrepancies from “good behavior” are detected. A certificate is revoked when a threshold number of accusations have been posted against one node. Any node is only allowed to post a single accusation against any other node.

The scheme offers limited robustness to varying network connectivity. The nodes are assumed to maintain both a common view of the number of nodes in the network and the behavior of these nodes, which is a strong requirement in ad hoc networks. Furthermore, a new node may lead to network congestion as the other nodes are supposed to send their certificates and profile tables (listing their view of earlier accusations and revocations) to newcomers.

In [15], Jungels, Raya and Hubaux suggest a revocation scheme for vehicular ad hoc networks (VANETs). VANETs differ from MANETs in the sense that the nodes move
along roads. The nodes communicate vehicle-to-vehicle or via base stations and fixed infrastructure available along the road. In each vehicle, the private-public key pairs are stored in a tamper-proof device. On revocations, the CA residing in the fixed net sends a key-erase message to the tamper-proof device. If the device does not confirm the erasure of all its keys, the CA warn other neighbors by a revocation list sent through a side channel, e.g., a *FM channel*. The key erasure and warning messages demand infrastructure. If the vehicles are out of range of a base station, the nodes warn each other through accusations in a way similar to the accusation based method of Crépeau and Davis [6]. However, Jungels, Raya and Hubaux’s scheme increase the robustness to packet losses as accusations are repeated periodically as long as the neighbors continue to receive suspicious messages. The accusations are forwarded to the CA as soon as a connection to the CA is detected. An accusation based scheme is also suggested in [7].

Morogan and Muftic [17] suggest the nodes of mobile ad hoc networks fetch CRLs when they are online on the Internet, or receive CRLs from nodes with fresher CRLs than their own.

To summarize, the proposals rely on the ad hoc nodes’ exchange of accusations or CRLs and downloads from trusted entities during periods with connection to the fixed network.

III Requirements

*Security:* The revocation information must be distributed in a manner that enables the recipient to verify its integrity, authenticity and freshness. The revocation information must reach the nodes in a timely manner.

*Robustness:* The distribution of revocation information must be robust both to packet losses and nodes exhibiting Byzantine behavior. In operational scenarios such as emergency and rescue operations, availability is a number one concern. No false revocations should ever occur.

*Simplicity:* Simplicity is an intuitive and overall design criterion. Computational complexity should be localized to the less constrained nodes. The decision to revoke a node may be made by an operator, but the distribution of revocation information should not involve human interaction.
Scalability: The revocation scheme should scale well enough to handle the expected number of nodes, security domains, and revocations. The revocation information must be distributed in a bandwidth efficient manner. Redundant and irrelevant revocation information should be minimized.

IV The MRL Scheme

A. Assumptions

A MRL (MANET revocation list) contains revocation information concerning nodes in a specific MANET. The term revocation list (RL) is used instead of certificate revocation list (CRL) as the method can be used both for traditional certificate based public keys as well as for identity-based schemes.

A central trusted entity is assumed in each security domain. This entity issues certificates or private keys, and is responsible for revocations. The keys are linked to long term IDs known by the routing protocol, e.g., OLSR main addresses [4], –either through certificates or by using the IDs as public keys. Each node may still use multiple and temporary identifiers. That is, only pre-defined nodes will expectedly be authorized to join protected ad hoc networks for emergency and rescue operations. The general assumption that the IP addresses should be obtained as the node enters the network, may not apply. Home addresses of Mobile IP may be utilized. This is described in [11].

A full RL is issued on a regular basis. Full RLs are transferred to the nodes prior to MANET participation. During MANET operation, MRLs are set up and distributed with the aid of trusted MANET gateways. The MRLs will expectedly be short. They are assumed to fit into a single routing protocol packet/message. MRLs could also be distributed over more packets/messages. However, the keys are used to bootstrap the network service. Long MRLs imply loss of network connectivity, which makes MRL distribution redundant.

The gateways are assumed to be more protected and less resource constrained, e.g., mobile base stations, than ordinary MANET nodes. Specific keys and IDs enable them to act as trusted gateways. The MANET may include nodes from a single security domain or from multiple security domains. The central trusted entity is assumed to possess computational and communicational resources large enough to handle the
expected number of concurrent MANET instantiations. It is also assumed to be accessible from the trusted MANET gateways when needed. A revocation may be a result of detected malicious behavior, a report of a lost unit or other.

B. Protocol Outline

1) MRL: central trusted entity – trusted MANET gateway

The protocol is shown in Fig.1. The notation “A->B: MSG NAME:” means that ID A sends a message named MSG NAME to ID B, and \(\{msg\}_{\text{signA}}\) means a message msg plus A’s signature on msg. In the figure, \(A\) refers to the trusted MANET gateway and \(B\) to the central trusted entity.

The MANET gateway initiates the setup of a MRL with a \textit{MRL initialization request (MREQ)} message (step 1 in Fig.1). The central trusted entity responds with a \textit{MRL-initialization proceed (MPRO)} message. The MPRO contains a fresh sequence number from the central trusted entity (SeqB1) and a receipt of the sequence number from the gateway (SeqA1). Old messages and messages where the signature check fails are discarded.

The MANET gateway proceeds with a \textit{Report & refresh (RR)} message that lists the IDs detected in the MANET. The central trusted entity registers the reported nodes and returns a MRL in step 4. The MRL lists the revoked IDs or revoked certificate serial numbers among the reported nodes. It may contain zero or more entries. The version number (RL no) is incremented only when new items are added to the list.

After the MRL has been established, the periodic RR messages serve the dual purpose of reporting new nodes discovered (if any) and informing the central trusted entity of the continued existence of the MANET. A MRL is returned in response to each RR message. Reported IDs need not be repeated. If a RR message has not been received within a given period of time, the central trusted entity considers the MANET terminated. If continued service is demanded, and the gateway has not received a MRL in response to a given number of RRs, it reverts to step 1.

A MANET may have more than one gateway to the external network. The central trusted entity maintains a separate MRL for each gateway, no matter whether the gateways report the same set of MANET nodes or not. The MANET gateways run one
instance of the protocol for each involved security domain, and receive a MRL from each involved central trusted entity. (The central trusted entities must exchange information about their trusted gateways, and issue cross-certificates or mutually signed system parameters.)

MRL initialization messages (initial handshake)
1. A→B: **MRL INIT. REQUEST (MREQ):**
   \{MREQ, A, SeqA1, B\}_signA
2. B→A: **MRL INIT. PROCEED (MPRO):**
   \{MPRO, B, SeqB1, A, SeqA1\}_signB

MRL maintenance messages (periodical)
3. A→B: **REPORT & REFRESH (RR):**
   \{RR, A, SeqA2, B, SeqB1, #IDs, IDs\}_signA
4. B→A: **MRL (MRL):**
   \{MRL, B, SeqB2, A, SeqA2, RL no, #revoked, revoked IDs\}_signB

General message format (some fields may be empty):
Message type, Source ID, Source sequence number, Destination ID, Last received sequence number from destination, List version number, # IDs on list, List of IDs, Source’s signature.
("#" = “The number of”)

Fig. 1 The MRL protocol between the trusted MANET Gateway (A) and the central trusted entity (B)

A→MANET nodes:
\{A, Message sequence number, [Ordinary routing message body], MRL \}_signA

Fig.2 MRL distribution in the MANET

2) **MRL: trusted MANET gateway – MANET nodes**

The MANET gateways distribute the MRLs as a separate routing message/packet or appended to an existing flooded routing message type. The protocol is illustrated in Fig.2. The MRL scheme assumes that the routing messages are protected by a cryptographic signature covering both the MRL as well as all other non-mutable fields of the routing message. The “ordinary routing message body” is only included if the MRL is appended to another routing message. Only non-empty MRLs are sent into the MANET. The latest non-empty MRL is flooded periodically. If multiple security domains are involved, more than one MRL may be included by the gateway.
3) Message processing in the MANET nodes

The revocation information distributed with the routing protocol needs only be checked when the first MRL is received and when the RL number indicates a change in the list. Routing messages from revoked nodes are discarded.

4) Autonomous ad hoc networks

When no trusted MANET gateway with access to the central trusted entity exists, revocation is precluded. That is, each node at all times monitors the behavior of its 1-hop neighbors. Nodes acting suspiciously are not accepted as symmetric neighbors. Suspicious behavior include actions such as forwarding modified messages, and announcing that it will forward traffic on other nodes’ behalf without doing so.

V The MRL Scheme with various routing protocols

**OLSR:** In the optimized link state routing protocol (OLSR) [2] [4], all nodes maintain a view of the topology of the entire ad hoc network. Topology information, including gateway announcements, is distributed periodically with the aid of Multipoint Relay (MPR) nodes. Each node selects MPRs from its 1-hop neighbors in such a way that all 2-hop neighbors are covered by at least one MPR. Only nodes that are chosen as MPRs forward routing information. Each node is uniquely identified by its main address that is included in all messages.

OLSRv1 [4] proposes that extensions to the protocol are implemented as new message types. OLSRv2 [2] adopts the generalized MANET packet and message formats specified in [3], and suggests extensions are implemented as new message types or type-length-value fields (TLVs) added to existing message types. For bandwidth efficiency, it is beneficial to include the MRL in existing routing messages. Additional packet transmissions and extra message signature validations are avoided.

OLSRv1 [4] specifies four message types: HELLO, Topology Control (TC), Multiple Interface Declaration (MID) and Host and Network Association (HNA). The HELLO messages are used for local link sensing, neighbor detection and MPR selection, and are not forwarded by the MPRs. The other message types are. TC messages convey topology
information, and are only sent by MPR nodes. The gateway may not be chosen as MPR. Nodes with multiple OLSR addresses send MID messages to map these interface addresses to the main address. The gateway may not have multiple OLSR interfaces. As the revocation information is distributed through gateways, HNA messages announcing non-OLSR interfaces are a natural candidate for the inclusion of MRLs. OLSRv2 only specifies HELLO and TC messages. Gateway announcements are included as TLVs in TC messages. Hence, gateways emit TC messages, and MRLs can be added as TLVs to these.

The inclusion of MRLs as TLVs in OLSRv1 HNA messages is sketched in Fig.3. The MRL scheme assumes that the routing messages are protected end-to-end by a message authentication TLV, e.g., using the scheme in [10]. Backward compatibility would require a new message type for MRLs. However, the inclusion of signature TLVs is also not backward compatible. According to the specifications in [4] signatures should also be distributed in separate messages. But, in [23] it is shown that end-to-end signatures in separate messages scale badly. And backward compatibility may not be important in operational scenarios such as emergency and rescue operations where only pre-defined nodes are allowed to join.

A potential problem is that only MPR selectors (nodes that have chosen this node as their MPR) are announced in the TC messages. Nodes that do not choose any MPRs could be accepted as symmetric neighbors, without having their existence exposed outside the scope of HELLO messages. However, in accordance with the OLSR specifications [4], the nodes can be set up to report all links. This will inhibit “hidden members”. Alternatively, the MPRs can report links to nodes that appears not to be included in any other TC messages.

**OSPF MANET:** OSPF MANET refers to the adaptation of the Open shortest path first (OSPF) [5] [18] routing protocol to ad hoc networks. Opposing OLSR, OSPF MANET provides reliable flooding. The proposed OSPF MANET solutions; **OSPF overlapping relays** (OSPF-OR) [1] and **OSPF MANET designated router** (OSPF-MDR) [20], differ in the way they optimize the flooding of routing information. OSPF-OR nodes choose overlapping relays (ORs) parallel to the MPRs in OLSR. Non-OR nodes act as backup ORs that retransmit the routing messages if the ordinary OR fails to do so. OSPF-MDR use **MANET designated routers** (MDRs) to flood routing information. The
The decision to become a MDR or a backup designated router (BMDR) is made by the nodes themselves. All routing messages are acknowledged.

The nodes are uniquely identified by their router ID included in each routing packet. All routers within an OSPF area have a consistent link-state database. The link-state database contains a collection of LSAs (link-state advertisements) that describes the OSPF routing domain. The topology information is disseminated through link state update (LSU) packets containing link-state advertisements (LSAs). Gateways to other areas (border area router) run different instance of the routing protocol for each area. Only summary information from one area is distributed into another.

### OLSR packet and message formats

<table>
<thead>
<tr>
<th>0-7</th>
<th>8-15</th>
<th>16-23</th>
<th>24-31</th>
</tr>
</thead>
<tbody>
<tr>
<td>Packet Length</td>
<td>Packet Sequence</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Packet header:**

<table>
<thead>
<tr>
<th>163</th>
<th>164</th>
<th>165</th>
<th>166</th>
</tr>
</thead>
<tbody>
<tr>
<td>Msg Type</td>
<td>Vtime</td>
<td>Message Size</td>
<td>Originator Address</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>167</th>
<th>168</th>
<th>169</th>
<th>170</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTL</td>
<td>HopCnt</td>
<td>Msg Seq.Number</td>
<td>Message authentication TLV</td>
</tr>
</tbody>
</table>

**Message header:**

<table>
<thead>
<tr>
<th>171</th>
<th>172</th>
<th>173</th>
<th>174</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRL TLV</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**HNA message:**

<table>
<thead>
<tr>
<th>175</th>
<th>176</th>
<th>177</th>
<th>178</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network address</td>
<td>Network address</td>
<td>Netmask</td>
<td>...</td>
</tr>
</tbody>
</table>

**Signature covers the entire message excluding mutable fields including the MRL TLV**

**Covered by revocation entity's signature**

---

Fig.3 OLSRv1 packet and message formats

We assume the MANET is one OSPF area. The area border routers announce their gateway capabilities through LSU packets containing inter-area-prefix LSAs and inter-area-router LSAs. We assume that the MRLs are included in LSU from trusted MANET area border routers. The other OSPF packet types; HELLO, database description (DD), link-state request, and link-state acknowledgement are not forwarded outside the 1-hop neighborhood.

The OSPF-OR and OSPF-MDR extensions to the OSPFv3 packet format is shown in Fig.4. MANET specific information is carried in a link local signaling (LLS) data block attached to HELLO and DD packets. OSPF-OR in addition specifies a new LSA type – link LSA that is used to distribute information about 2-hop neighbors. Each LLS data block may contain several TLVs. LLS-incapable routers will not consider extra data that
follows after the packet. Thus, the attached LLS data block renders the extended packet format compatible with standard OSPF [26]. However, if each MANET is a separate OSPF-area, and only predefined nodes are allowed to join, compatibility with OSPFv3 may not be a strict requirement.

We suggest MRLs distributed in TLVs in a LLS data block attached to the LSU packets. The scope of the LLS thus becomes MANET-local rather than link-local. OSPF-OR [1] assumes LLS data blocks are appended to HELLO or DD packets used in the synchronization of the link-state databases of adjacent nodes. But LLS data blocks could also be attached to other packet types. We assume end-to-end protection of the routing information is implemented with signatures in TLVs in the LLS data block. It requires LSU packets are flooded as is. (OSPF assumes LSAs are flooded. The LSU encapsulating the LSAs could be modified hop by hop.)

Whereas OSPFv2 [18] specifies authentication as an integral part of the routing protocol, OSPFv3 [5] relies on IPsec. OSPF-OR suggests the ability to connect to the MANET is controlled by layer 2 security mechanisms such as IEEE802.11i. The MRL scheme can be used no matter which layer protects the routing messages. That is, at least as long as an asymmetric scheme is used and the revoked keys can be linked to IDs known to the routing protocol. IPsec and IEEE802.11i by default use symmetric keys. Unilateral authentication of ad hoc routing messages may imply a modification to the existing standards.

![Fig.4 OSPF MANET packet extension](image)

**AODV:** In the ad hoc on demand distance vector (AODV) [21] routing protocol, routes are discovered on demand by flooding a route request (RREQ) into the network. The RREQ carries the IP addresses of the originator as well as the destination. The destination (or intermediate nodes with a valid route to the destination) unicasts a route reply (RREP) along the path that the RREQ was received. The AODV specification [21] allows extensions in the form of type-length-value (TLVs) appended to RREQs and
RREPs. We assume that all routing messages are integrity protected and authenticated by digital or identity-based signatures.

The gateways will normally not have a complete map of all nodes in the MANET. They can listen to the routing messages that are flooded in the network, and report to the central trusted entity as IDs are discovered. However, the destination will return a RREP along the path it received the RREQ. If the destination lies between the source and the gateway, the gateway may not receive any of the messages. The AODV expanding ring search, i.e., the flooding scope of RREQs is gradually increased until a route reply is received, makes it even less likely. Besides, intermediate nodes in a path will not be announced. Inclusion of the MRL scheme demands a modification of the protocol: all nodes on an active path must periodically send a RREQ searching a gateway. The “D flag” must be set to avoid intermediate nodes with a valid route answers the request. The gateway checks the signature of the RREQ before the node is reported. If a node on an active path does not hear such RREQs originated from one of the precursors within a given period of time, the precursor is reported in a TLV appended to its next gateway RREQ or blacklisted (excluded).

The MRLs are added to RREQs sent by the gateway to a predefined revocation information address. All nodes processes and forwards the RREQ, but do not return any RREPs.

VI Analysis

A. Security evaluation

A Dolev-Yao threat model [9] is assumed, i.e. the adversary may eavesdrop on any message, modify and replay messages, and forge source and destination addresses. But the adversary cannot produce a valid signature on other nodes’ behalf or decrypt a message that has been encrypted with a key she does not possess. That is, for the ad hoc network the Dolev-Yao threat model must be modified to encounter that one or a few nodes may be compromised. This is as ad hoc nodes are expected to be out in the field. The probability of compromised nodes will expectedly be higher than in a fixed network.
1) Security of the protocol between the MANET gateway and the central trusted entity:

The first three steps of the protocol in Fig.1 are parallel to the three-pass mutual authentication mechanism specified in ISO 9798-3[13]. That is, the random numbers in the ISO standard are replaced by sequence numbers in the MRL protocol. Mutual authentication is in both cases obtained after the first three steps have completed. However, in order to continue to evaluate the freshness of the next messages, the MRL protocol requires that each message is linked to the preceding one. Differently from the random numbers of the ISO standard, the incrementing sequence numbers links the next message to the previous one beyond the first three steps. The sequence numbers are not reused during the lifetime of the signature keys. Each message is thus uniquely identifiable.

The signature on the initial MREQ message proves the integrity and authenticity of origin, but not the freshness of the request. The central trusted entity can judge from the sequence number whether it has responded to this request earlier or not. But if not, at this point it cannot decide whether the request is fresh or not. The MANET gateway’s sequence number (SeqA1) returned in the MPRO message enables the MANET gateway to decide which MREQ this MPRO is a response to. Similarly, the central trusted entity is able to evaluate the freshness when it receives the receipt of its sequence number in the RR message in step 3. The sequence numbers prevents that MRLs are set up on the basis of replays. The sequence number receipt in the MRL serves to acknowledge the list of nodes reported in the previous RR. It also enables the MANET gateway to detect MRL replays.

The MANET gateway can scale the **timeliness** of MRLs to the desired level by adjusting the frequency of RR messages.

2) The security of the protocol between the MANET gateways and the MANET nodes:

In Fig.1 the signature of the MANET gateway ensures the message integrity and authenticity of origin of the routing message that encapsulates the MRL. The central trusted entity's signature ensures the integrity and authenticity of the MRL. The
inclusion of the ID of the gateway in the signed part of the MRL prevents others from posing as a trusted gateway by including an old list issued to another gateway.

The sequence numbers and the \textit{RL no} included in the MRL signed by the central trusted entity indicate the freshness of the information. Higher numbers imply fresher information. However, these numbers do not uniquely link the MRL to this specific MANET instantiation. In theory, the trusted gateway could replay MRLs from an earlier MANET instantiation. A way to reduce this problem would be to include the version number of the latest \textit{full revocation list} in the MRL. This would reduce the window of opportunity for successful replays of MRLs. A lower version number in the MRL than the version number of the latest revocation list held by the nodes would imply a replay. However, under the assumption that certificates or IDs are never revalidated after their revocation (at least not during the lifetime of the public key/system parameters of the central trusted entity), the replay of old MRLs does no other harm than wasting bandwidth.

The nodes must trust the MANET gateway to correctly report all nodes in the MANET to the central trusted entity. With the protocols in Fig.1 and Fig.2 the gateway could undetectably just report a subset of the nodes. The inclusion of the list of reported nodes in the MRL message would mitigate this. However, it would also increase the bandwidth cost. Besides, the gateway could still refuse to forward the MRL. We suggest redundant gateways as a more practical and robust solution. As long as at least one gateway behaves properly, the relevant revocation information will be distributed into the MANET.

A malicious insider may not want to forward revocation information about itself. But other nodes that behave according to the protocol will. In OSPF MANET the backup MDRs or non-OR nodes will forward the message if the ordinary MDR/OR fails to do so. In OLSR, if one MPR does not forward the MRL, the node will still receive it from other symmetric neighbors that retransmit the message. Though, the network may contain sparsely connected areas. If a node represents the only connection from the gateway to the rest of network, this node could hold back revocation information from propagating into the part of the network that it “controls” -at least temporarily. Node mobility, changing topology and the periodic (re)transmissions of MRLs reduce the problem.
B. Robustness and simplicity

The scheme is simple and robust in the sense that there is no need to synchronize the revocation lists, even if not all ad hoc nodes have the latest full RL. If a revoked node is included in the network, it will appear in the next MRL, and be excluded. There is no need for the gateways to synchronize their lists of nodes, and the central trusted entity needs not harmonize the various MRLs. The scheme is robust to gateway failures in the sense that MRLs are deleted automatically if the central trusted entity has not receive any RRs within a given timeout period. The scheme is robust to packet losses. Non-empty MRLs are repeated in the periodical routing messages from the gateways. Bandwidth consumption is optimized as the MRLs only include information concerning the nodes in the specific MANET. Also, most of the MRL administration is off-loaded to gateways and central trusted entities that are expected to be less energy-constrained than the battery powered ad hoc nodes.

No revocation information will be distributed unless a trusted gateway with access to the fixed net exists. A remedy could be to empower more protected and trusted ad hoc nodes to deny access, possibly implemented by a threshold scheme [8] (a threshold number of trusted nodes must co-operate to exclude a node). But it also adds complexity and increases bandwidth cost.

C. Scalability/Performance

The MRL scheme is here compared with CRLs and $\Delta$-CRLs distributed through the gateways rather than the state-of-the-art accusation-based and CRL-exchange methods surveyed in section II. This is as we consider CRLs and $\Delta$-CRLs to be more likely alternatives for our operational scenario. Accusation based schemes may lead to false revocations.

1) The number of revocations

Key changes represent a large administrative cost. Changes of keys used to bootstrap the network service also represent a threat to network availability. Hence, in practical scenarios keys can be expected to last for more than one MANET instantiation. With increasing lifetimes for the keys, the anticipated number of revocations will also increase. The number of revocations in a specific MANET will depend on the size of the MANET and the lifetime of the network. The probability of a revocation during a
MANET operation $p_{\text{revMANET}}$ can be estimated from the revocation probability $p_{\text{rev}}$ of the involved security domains and the lifetime of the MANET $t_{\text{MANET}}$ to the lifetime of the key/certificate $t_{\text{keylife}}$:

$$p_{\text{revMANET}} = p_{\text{rev}} * \frac{t_{\text{MANET}}}{t_{\text{keylife}}} \quad (2)$$

With $N_{\text{tot}}$ representing the total size of the security domains, the total number of revocations in the involved security domains during a MANET instantiation, $N_{\text{rev}}$, equals

$$N_{\text{rev}} = N_{\text{tot}} * p_{\text{revMANET}} \quad (3)$$

However, with $N_{\text{MANET}}$ nodes in the MANET, only $N_{\text{MANET}}/N_{\text{tot}}$ of these will expectedly concern nodes in this MANET. The number of revocations in a specific MANET during its lifetime $N_{\text{revMANET}}$ is thus:

$$N_{\text{revMANET}} = N_{\text{MANET}} * p_{\text{revMANET}} \quad (4)$$

To exemplify this: assume a uniform distribution of revocations with $p_{\text{rev}}=10\%$, $N_{\text{tot}}=100000$, $N_{\text{MANET}}=100$, $t_{\text{MANET}}=1\text{ day}$, $t_{\text{keylife}}=365\text{ days}$. The expected average number of revocations within a specific MANET will then be $N_{\text{revMANET}} < 1$. The expected total number of revocations within the involved security domains during the same period equals $N_{\text{rev}} \approx 28$. With the assumed values, a $\Delta$-CRL issued one week after the latest full RL would contain around 200 entries. A full RL also comprises previously revoked nodes.

2) Overhead

Overhead between the trusted gateways and the central trusted entity depend on the RR frequency, which can be scaled by the gateways. The sizes of the messages are confined by the number of nodes in the MANET and the signatures and message header information inserted by the gateway and the central trusted entity. The signatures will expectedly contribute most. Redundant revocation information is reduced as origination of MRL messages is limited to trusted MANET gateways, and only non-empty MRLs are sent into the network.

The scalability of the MRL scheme is illustrated in Fig.5. The figure shows an estimate for the expected average channel utilization under the assumption that the revocation information is included in OLSRv1 HNA messages. Channel utilization refers to the fraction of a second that the channel is occupied with OLSR traffic. The figure shows the
cost of including a MRL, CRL or Δ-CRL in protected OLSR messages. Standard OLSR is included for reference. Topology information and calculation methods are adopted from [23], and were obtained through calculations and simulations with the ns-2 network simulator. The calculations assume IEEE 802.11b with CSMA/CS-DSSS for the lower layers, and OLSR traffic broadcast data rate of 1Mbps. UDP, IP and 802.11 MAC headers add 8+20+34 bytes to each OLSR packet, respectively.

A collision and error free channel is assumed. Additional 802.11b delay is 552us [14] per MAC frame. The calculations are based on the formats in Fig.3.

Protected OLSR refers to the inclusion of a 512 bit signature in all routing messages. The calculations assume one gateway that emits one MRL, CRL or Δ-CRL appended to its protected OLSR HNA messages. A CRL of 2000 entries, a Δ-CRL with 200 entries, and a MRL with 1 entry is assumed. Each entry is a 32 bit ID. The MRL, Δ-CRL and CRL also include a 2048 bit signature from the central trusted entity.

The figure indicates a significant bandwidth cost for the distribution of CRLs. MRLs or Δ-CRL contributes considerably less to the channel utilization. With multiple security domains and a CRL or Δ-CRL from each domain, the bandwidth savings obtained with the MRL scheme compared to CRL and Δ-CRL will be even greater. The same holds true if more gateways insert revocation information.

**D. Computational cost**

Keys of central trusted entities are normally longer than those of ordinary nodes. The signature verification cost increases accordingly. This is optimized in that the MANET nodes only have to verify the MRL signature when a new MRL is flooded into the
network. The number of changes will usually be much lower than the number of nodes in
the network.

VII Conclusions and Future Work

In the MRL scheme, only revocation information that explicitly concerns the nodes in
a specific MANET is disseminated into the ad hoc network. The overhead depends on
the number of nodes in the ad hoc network rather than the total number of nodes in the
involved security domain(s). This makes the MRL scheme scale well.

The scheme requires trusted MANET gateways can detect which nodes are in the
network. This comes intrinsically with proactive routing protocols. We also described the
implementation of the MRL scheme with reactive routing protocols. The impact on the
routing protocol performance is a topic for further analysis. Further work is also needed
to provide a formal security analysis of the protocol between the trusted MANET
gateway and the MANET nodes. The formal security analysis requires a method that
takes the nature of the ad hoc environment and the characteristics of the specific routing
protocol into consideration.

The MRL scheme has been designed for the revocation of certificates or IDs used to
protect the network layer routing information. The technique could also be utilized for
application data. This is another topic for further research. Whereas keys used to protect
the ad hoc routing information have a MANET internal scope, keys for application data
may have a global scope.

References

Internet-draft, April, 2005


[10] A.M. Hegland, P. Spilling, L. Nilsen, and Ø. Kure, ”Hybrid Protection of OLSR,” *Workshop on Cryptography for Ad hoc Networks (WCAN’06)*


